TEMPORAL INFORMATION PROCESSING AS A FOUNDATION OF SPEECH PERCEPTION

Thesis
presented to the Faculty of Arts
of
the University of Zurich

for the degree of Doctor of Philosophy
by

Tino Zähle

Accepted in the summer semester 2007

on the recommendation of

Prof. Lutz Jäncke
PD. Valentine Marcar

Zurich 2007

Contents

Acknowledgements **IV**

Summary **V**

Zusammenfassung **IX**

1. Introduction **1**

1.1 Present models of functional lateralization of basic auditory information processing 2

1.2 Clinical research 4

1.3 Behavioural research on the perception of temporal cues 5
1.3.1 The Perception of Voice-onset-time (VOT) 5
1.3.2 The Perception of Noise-onset-time (NOT) 6

1.4 Evidences from neuroimaging 7

1.5 Methods: Functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) 10
1.5.1 Functional magnetic resonance imaging (fMRI) 11
1.5.1.1 Basic principles of MRI scanning 11
1.5.1.2 Methods of analysis 17

1.5.2 Electroencephalography (EEG) 21
1.5.2.1 Recording of human scalp EEG 21
1.5.2.2 The electrophysiological basis of EEG 22
1.5.2.3 Event-related potentials (ERPs) 24

1.6 Reference List 26

2. Methodological examinations **33**

2.1 *Study A:* Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task 35
Abstract 35
Introduction 37
Materials and Methods 40
Results 46
Discussion 51
Reference List 56

2.2 *Study B:* Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks 63
Abstract 64

Introduction 65
Materials and Methods 68
Results 74
Discussion 81
Reference List 84

3. Own previous related work as background and originator for further investigations 87

3.1 *Study C:*
Evidence for rapid auditory perception as the foundation of speech processing – a sparse temporal sampling fMRI study 88
Abstract 89
Introduction 90
Methods 91
Results 99
Discussion 104
Reference List 111

4. Investigation 116

4.1 *Study D:*
Spectral-temporal processing during speech perception involves left posterior auditory cortex 119
Abstract 120
Introduction 121
Methods 122
Results 125
Discussion 127
Reference List 131

4.2 *Study E:*
The neural correlate of speech rhythm as evidenced by meter processing: an fmri study. 133
Abstract 134
Introduction 135
Material and Methods 138
Results 144
Discussion 148
Reference List 155

4.3 *Study F:*
Evidence for a general segmental processing device in the human auditory dorsal stream 164
Abstract 165
Introduction 166
Material and Methods 168
Results 173
Discussion 179
Reference List 187

4.4 *Study G:*
Electrical brain imaging reveals overlapping activity elicited by temporal cue processing in speech and non-speech sounds 193
Abstract 194
Introduction 195
Material and Methods 197
Results 202
Discussion 208
Reference List 212

5. Concluding Discussion **217**

5.1 Methodological examinations 217

5.2 Investigation 219

5.3 Reference List 224

6. Curriculum vitae **225**

Acknowledgements

There are many people I wish to thank for helping and supporting me in various phases of my PhD. In the following lines some of them are gratefully acknowledged. However, I am aware of the fact that there are many more and these words cannot express the gratitude and respect I feel for all of those.

Firstly, I like to take this opportunity to thank my supervisor Prof. Dr. Lutz Jäncke at the University of Zurich who believed in me since my early days as a student at the psychology department in Magdeburg. Thank you very much for your supportive scientific supervision and your tolerance in giving me this great topical freedom.

Furthermore, I am most grateful to Dr. Martin Meyer. During the time of my PhD his advice was always competent and professional. Martin, thank you very much for this very fruitful and instructive time. I hope to continue collaborating with you.

I like to express further greatest thanks for help and encouragement to Prof. Dr. Fred Mast, who was the first who offered me a warm welcome in his group in Zurich. Fred, thank you for your support and guidance during the whole time of my PhD.

A special thank goes to my co-workers at the University of Zurich, who created such a pleasant working environment and fruitful scientific stimulation. Special thanks go to the *doctores* Janek Lobmaier, Peter Klaver, and Kai Lutz, as well as to Stephanie Rotzer, Stephan Elmer, Eveline Geiser, Mathias Oechslin and Pascale Sandmann.

Last but not least I am deeply grateful and indebted to my friends and colleagues Dr. Nico Bunzeck and Dr. Torsten Wüstenberg for there always amicable advices.

Finally, my deepest thanks go to my family. I am particularly indebted to my parents for their never-ending encouragement and ongoing support. Very special thanks go to Inga Bonath. Thank you for your love, friendship and succour.

Summary

The classical brain-language model derived from the work of Broca, Wernicke, and Lichtheim, which favors the existence of 'speech centers' in the left hemisphere, has increasingly been challenged in the past few years. Recent models of speech comprehension describe both the left and the right hemisphere as governing signal analysis during auditory language processing (Boemio *et al.,* 2005; Hickok and Poeppel, 2000; Hickok and Poeppel, 2004; Poeppel, 2003) with the left hemisphere preferentially extracting rapid information from short integration windows while the right hemisphere is more adept at extracting slow information from long integration windows.

In the present work a series of fMRI and EEG studies was performed to test the models' predictions on sub-lexical speech processing by investigating neurophysiological responses to temporally and spectrally modulated speech and non-speech sounds. In particular, this work specifically investigated the neural underpinnings of functional lateralization in speech perception. Brain responses in the peri-auditory regions of auditory speech and non-speech stimuli (indexed by functional magnetic resonance imaging (fMRI) and auditory evoked potentials (AEP)) are assumed to vary systematically as a function of acoustic modulations in the sound signal. The fMRI/EEG studies of the present work aim at uncovering the basic processing principles in the auditory domain, which forms the basis for the interplay between acoustic and phonological processing during speech processing. Furthermore, to provide a methodological base for studying basic auditory processing using fMRI, the implementation and evaluation of "silent" fMRI acquisition protocols was specifically addressed.

The first goal of the present work was the implementation and evaluation of "clustered-sparse temporal acquisitions (CTA)" protocols in auditory fMRI research.

In **study A,** entitled "*Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task*", we investigated the effects of ambient scanner noise bound to continuous fMRI on hemodynamic responses to spoken sentences by comparing conventional noisy fMRI with a silent clustered acquisition scheme. In the results we observed that auditory processing in a silent fMRI paradigm brought on stronger functional responses of the bilateral supratemporal plane accommodating the entire auditory cortex, whereas the presentation of the identical auditory stimulus material in the context of a traditional continuous acquisition more strongly recruited the bilateral insulae. This investigation clearly demonstrated that "clustered-sparse temporal acquisitions" protocols can successfully be applied to map functional responses to auditory sentence stimuli in a silent environment.

In **study B,** entitled "*Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks*", we evaluated a clustered (CTA) and a sparse temporal acquisition protocol (STA). The comparison of the two fMRI acquisition schemes revealed that, for the left and right temporal cortex, the auditory stimulation related number of suprathreshold voxels was larger and the maximum t-value was significantly higher for the CTA- relative to the STA scheme. To summarize, this study shows that to use the full benefit of "silent" fMRI it is advantageous to acquire several sample points per trial regardless of the stimulus duration. Furthermore, data demonstrate that the acquisition of a cluster of subsequent imaging volumes along the T1-decay can successfully be applied in the context of auditory fMRI research.

The second goal of the thesis aimed to uncover the relation between acoustic and phonological processing during speech perception

In **study D,** entitled "*Spectral-temporal processing during speech perception involves left posterior auditory cortex*", we investigated the neural underpinnings of spectral-temporal integration during speech perception.

The study consisted of a behavioral pilot study and a follow-up fMRI-experiment to elucidate the neural underpinnings of basic linguistic processing. We presented participants with sine-wave analogues that could be either perceived as non-speech auditory forms (naïve condition) or, after instruction and a short practice, as speech (informed condition). Behavioral results revealed a difference in the processing mode; spectral-temporal integration occurred during speech perception, but not when stimuli were perceived as non-speech. In terms of neuroimaging we observed an activation increase in left posterior primary and secondary auditory cortex, namely Heschl's gyrus and planum temporale encroaching onto the superior temporal sulcus, reflecting a shift from auditory to speech perception. This finding demonstrates that the left posterior superior temporal lobe is essential for spectral-temporal processing during speech perception.

In **study E,** entitled "*The neural correlate of speech rhythm as evidenced by meter processing: an fmri study*", we investigated the neural correlates of rhythm processing in speech perception using a comparison of German pseudo-sentences spoken with an exaggerated (metric) or a conversational (non-metric) rhythm in the context of an auditory clustered-temporal-sampling fMRI experiment. For the perception of speech rhythm, suprasegmental cues such as syllable duration, syllable stress, or pause are most important. Therefore, the analysis of these acoustic suprasegmental speech characteristics mainly relies on the perception of temporal intervals. Thus the perception of speech rhythm incorporates an appropriate tool to investigate auditory acoustic temporal processing beyond the perception on the level of segmental speech characteristics such as consonants and vowels. Results of the study indicate a function of SMA and insula beyond motor timing and speak for a role of these brain areas in the perception of temporal intervals. Furthermore, the data demonstrate a specific task-related function of the right IFG in the processing of accent patterns and show right secondary auditory cortex involvement in the explicit perception of auditory cues and that activity in the right secondary auditory cortex can be modulated by top-down mechanisms.

In **study F,** entitled "*Evidence for a general segmental processing device in the human auditory dorsal stream*" we investigated the neurofunctional organization of sub-lexical auditory perception with specific respect to auditory spectro-temporal processing in speech and non-speech sounds. Participants were presented with verbal and nonverbal auditory stimuli with systematic variations of the spectral or temporal acoustic characteristics in the context of a sparse event-related fMRI study. Results of this investigation show a subdominant left hemisphere involvement during the processing of rapidly changing temporal characteristics regardless of the speechness of the sounds. In particular, when participants had to attend to subtle temporal modulations within speech and non-speech stimuli we observed significant responses in the parietal operculum and the frontal operculum on the left hemisphere. In contrast, we revealed no lateralization effects for spectral processing. When participants had to listen to changes in the frequency content of speech and non-speech stimuli, activations of the medial temporal gyrus und superior temporal sulcus bilaterally were explored.

In **study G,** entitled "*Electrical brain imaging reveals overlapping activity elicited by temporal cue processing in speech and non-speech sounds*", scalp AEPs in response to CV-syllables and non-speech analogues with varying VOT and noise-onset-time (NOT), respectively were recorded. This study investigated the neural coding of acoustic characteristics underlying speech perception. Results showed that the characteristic AEP waveform in response to consonant-vowel-syllables can be resembled by non-speech sounds with similar temporal characteristics, and that the primary auditory cortex is asymmetrically activated in favoring the left hemisphere independent of the stimulation used. Thus the study demonstrated that the early electrophysiological responses (up to 200 ms after stimulus onset) are not modulated by the linguistic content of the auditory signal, but greatly depend on the acoustic features of the processed sounds.

Zusammenfassung

Basierend auf den Ergebnissen aktueller empirischer Untersuchungen und modernen Theorien zur Verarbeitung von sprachlichen Reizen im auditorischen System wurden in dieser Arbeit vier Studien entwickelt und durchgeführt. Ziel dieser Experimente ist die detaillierte Untersuchung der Zusammenhänge zwischen akustischen und phonologischen Verarbeitungsprozessen im Rahmen der auditorischen Sprachwahrnehmung. Darüber hinaus bestand ein weiteres Ziel der vorliegenden Arbeit darin, ein neuentwickeltes kernspintomographisches Untersuchungsverfahren zu evaluieren und zu etablieren.

Die Studien A und B wurden entwickelt und durchgeführt, um ein neuartiges kernspintomographisches Untersuchungsverfahren, welches die Möglichkeit einer geräuscharmen Experimentaldurchführung bietet, das so genannte “clustered-sparse temporal acquisitions (CTA)”, zu evaluieren.

Die **Studie A** mit dem Titel “*Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task*” hatte zum Ziel, das CTA Protokoll mit einem klassischen fMRT Untersuchungsdesign mit starker Geräuschbelastung zu vergleichen. Die Ergebnisse dieser Untersuchung erbrachten einen klaren Vorteil für die Verwendung des CTA Protokolls. Die hämodynamischen Antworten im gesamten perisylvischen auditorischen Kortex waren erhöht und die Versuchspersonen waren in der Lage, die an sie gestellten Aufgaben schneller zu lösen, wenn die Datenaufnahme mit einem CTA Protokoll durchgeführt wurde.

In der **Studie B** mit dem Titel “*Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks*” wurde das CTA- mit einem bereits etablierten STA-Protokoll verglichen. Es zeigte sich, dass die statistische Analyse der mit dem CTA-Protokoll

erhobenen Daten eine größerer Anzahl an signifikant aktivierten Voxeln und höhere t-Werte hervor brachte, als die Analyse der mittels STA erhobenen Daten. Es zeigte sich erneut ein klarer Vorteil für die Verwendung eines CTA- Protokolls bei der Untersuchung auditorischer Kortexfunktionen.

Die Studien D-G wurden durchgeführt mit dem Ziel einer detaillierten Untersuchung der Zusammenhänge zwischen akustischen und phonologischen Verarbeitungsprozessen im Rahmen der auditorischen Sprachwahrnehmung.

Die **Studie D** mit dem Titel "*Spectral-temporal processing during speech perception involves left posterior auditory cortex*" hatte zum Ziel, die funktionell-neuroanatomischen Grundlagen der Verarbeitung spektraler und temporaler Aspekte der auditorischen Sprachwahrnehmung mittels funktioneller Kernspintomographie (fMRT) zu untersuchen. Die Ergebnisse dieser Untersuchung demonstrierten die entscheidende Rolle des linken posterioren gyrus temporalis superior (GTS) bei der integrativen Verarbeitung spektraler und temporaler Eigenschaften des auditorischen Sprachsignals.

Die **Studie E** mit dem Titel "*The neural correlate of speech rhythm as evidenced by meter processing: an fmri study*" untersuchte mittels fMRT die neuronalen Korrelate der Rhythmusverarbeitung im Rahmen der Sprachwahrnehmung. Die Ergebnisse dieser Untersuchung zeigten spezifische Aktivierungen des suplementär motorischen Areals (SMA) und der Insula bei der Verarbeitung von akustischen, suprasegmentalen auditorischen Informationen. Des Weiteren zeigte diese Studie eine spezifische aufmerksamkeitsabhängige Funktion des rechten gyrus frontalis inferior (GFI) bei der Sprachrhythmusverarbeitung.

In der **Studie F** mit dem Titel "*Probing a parameter-specific model of auditory lateralized processing by fMRI*" wurde mittels fMRT die funktionell-neuroanatomischen Organisation der sub-lexikalischen auditorische

Verarbeitung untersucht, und die spektralen und temporalen Wahrnehmungsprozesse systematisch manipuliert. Die Ergebnisse dieser Untersuchung zeigten eine spezifische Aktivierung eines linkshemisphärischen fronto-parietalen kortikalen Netzwerkes bei der Verarbeitung von sich schnell ändernden akustischen Informationen und demonstrierten wesentliche funktionelle Eigenschaften des dorsalen Verarbeitungspfades im Rahmen der auditorischen Sprachwahrnehmung.

Die **Studie G** mit dem Titel "*Electrical brain imaging reveals overlapping activity elicited by temporal cue processing in speech and non-speech sounds*" untersuchte mittels Elektroenzephalographie (EEG) die Verarbeitung von Silben mit variablen voice-onset-times (VOT) und die Verarbeitung von nicht-sprachlichen Stimuli mit analogen akustischen Eigenschaften (noise-onset-time, NOT). Diese Untersuchung konnte zeigen, dass die elektrophysiologischen Reaktionen in den ersten 200 ms nach dem Beginn eines auditorischen Stimulus nicht davon abhängen, ob es sich bei dem Signal um einen sprachlichen oder nicht-sprachlichen Stimulus handelte. Die Elektrophysiologische Reaktion ist ausschließlich moduliert durch die akustischen Eigenschaften der zu verarbeitenden Signale.

1. Introduction

The classical view of hemispheric specialization of language organization within the cortex goes back to observations on patients with brain lesions. In 1861, Broca reported a post-mortem study of a patient who had been impaired at articulating language. This patient had extensive damage in the fontal lobe of the left hemisphere. In 1874, Wernicke reported a post-mortem study of a patient with impaired speech comprehension and a lesion in the left posterior superior temporal cortex. Beginning with these lesion studies, an early model of lateralized speech processing was developed based on the idea that language was localized in structures which are functionally and structurally interconnected to accomplish the brain's language system (Lichtheim L., 1885). This model, referred to as the classical localizationist view was dominant through a long time of clinical research basically on aphasia.

Early studies regarding hemispheric specialization for speech perception using dichotic listening techniques found that volunteers exhibit a right ear advantage for most speech stimuli, indicating a greater left hemispheric processing (Bryden, 1982; Kimura, 1961; Kimura and Folb, 1968). Analogously, a left ear advantage was also observed for the identification of certain non-speech stimuli, including sounds, music or white noise (Kimura, 1964; King and Kimura, 1972).

These early findings provided further evidence for the view that language abilities are left lateralized and non-language abilities are lateralized to the right hemisphere.

Schwarz and Tallal (1980) suggested, that the right ear advantage might reflect phonetic as well as acoustic processing. They assumed that the left hemisphere is sensitive for rapidly changing acoustic cues in general. These rapidly changing acoustic cues are frequent characteristics of linguistic stimuli, but it is not the linguistic stimuli per se that accounts for left hemispheric processing. The authors based their assumption on the finding that a deficit in the temporal information processing within the

auditory system, rather than a deficit in the linguistic abilities, could account for phonological disorders observed in language-learning-impaired children (Schwartz and Tallal, 1980; Tallal and Gaab, 2006; Tallal and Newcombe, 1978). This hypothesis seems to be confirmed by the improvement of speech discrimination and language processing in these children after training of recognition of temporally modulated speech and non-speech sounds (Merzenich *et al.,* 1996; Tallal *et al.,* 1996). Furthermore, the combination of deficits in auditory processing of speech and non-speech sounds that included rapid changing acoustic cues has also been shown for adult aphasics (Steinbuchel and Poppel, 1993). Taken together, rapid temporal information processing seems to be a fundamental mechanism in perception of speech and non-speech stimuli within the auditory domain.

Recent research developments have been furthering our knowledge about the rapid temporal information processing within the auditory system in human. Main research landmarks are (i) the recent models describing the functional lateralization of temporal auditory information processing, (ii) evidence from clinical research, (iii) results from behavioral research investigating the perception of temporal cues in speech and non-speech analogues, and (iv) recent examination of acoustic processing within the human auditory system using modern brain imaging methods.

In the following sections these issues will be briefly described and the most important findings of these research fields will be addressed.

1.1 Present models of functional lateralization of basic auditory information processing

Based on neuropsychological observation on patients with acquired brain lesions and on the large body of neurophysiological studies, several theories concerning lateralized auditory processing have been postulated. It has been proposed that speech related left hemisphere lateralization arises as a result of the linguistic relevance and intelligibility of the input (Scott *et al.,* 2000). Alternatively, it has been argued that the hemispheric specialization exists as a result of asymmetries in basic auditory processing

(Efron, 1963; Tallal and Piercy, 1973a). More generally, Zatorre and colleges (Zatorre *et al.,* 2002a) proposed a model suggesting that the left superior temporal cortex is specialized for the analysis of temporal changes of auditory signals whereas the right temporal cortex is dedicated to spectral analysis. Likewise, the model proposed by Poeppel (Poeppel, 2001; Poeppel, 2003) argues that the speech signal is analyzed in separate time domains whereby the left hemisphere preferentially extracts information over shorter temporal integration windows and right hemisphere over longer windows. This asymmetric sampling in time (AST) model posits that the initial spectro-temporal representation is bilaterally symmetric; followed by a temporally asymmetric elaboration where the auditory signal is analyzed on two different time scales. Integrating these models of functional hemispheric lateralization, Hickok and Poeppel proposed a large-scale model of the functional anatomy of language (Hickok and Poeppel, 2000; Hickok and Poeppel, 2004). The model posits that the cortical processing system involved in speech perception at an early stage involves the superior temporal gyrus (STG) bilaterally and diverge afterwards into two processing streams, a ventral stream, which is involved in sound-to-meaning mapping and a dorsal stream, which is involved in sub-lexical speech processing. In this model, the ventral stream involves cortical regions in the vicinity of the temporo-parietal-occipital-junction. The dorsal stream projects towards the parietal lobe and to frontal regions. Likewise, Johnsrude and Scott (Scott and Johnsrude, 2003) proposed a neuroanatomical model of speech perception that postulates functionally distinct processing streams extending anteriorly and inferolaterally from primary auditory cortex. The anterior system involves the anterior superior temporal sulcus and ventrolateral and dorsolateral frontal cortex. The system is involved in mapping acoustic-phonetic cues onto lexical representations. The posterior system involves posterior auditory belt and parabelt, the posterior STS, parietal cortex and ventrolateral and dorsolateral frontal cortex. This system processes articulatory-gestural speech-relevant representations.

1.2 Clinical research

Neuropsychological findings have suggested that the fine temporal resolution in auditory perception is closely linked to the left hemisphere, specifically to areas that are associated with language processing (Tallal and Newcombe, 1978). Patients with acquired brain lesions and aphasia (Efron, 1963; Swisher and Hirsh, 1972), children with general language-learning disabilities (Tallal and Piercy, 1973b; Tallal and Stark, 1981), and children and adults with dyslexia (Tallal, 1980) have been shown to be impaired in the identification of temporal sequences of non-verbal stimuli. Furthermore, patients with left sided lesions and symptoms of a fluid aphasia exhibited strongly reduced performances in the auditory processing of temporal information (Robin *et al.,* 1990; Von Steinbuchel *et al.,* 1999). The ability of these patients in phoneme discrimination could be increased by training of the auditory order threshold (Mates *et al.,* 2001; Von Steinbuchel, 1995).

It has been shown that children with reading disabilities are deficient in phoneme perception which is reflected by inconsistent labelling of tokens in a Voice Onset Time (/ga/–/ka/) series (Tallal, 1980; Tallal *et al.,* 1993). At the same time these children are less consistent in labelling of Tone Onset Time tokens (Breier *et al.,* 2001), and exhibit poorer auditory order thresholds (Von Steinbuchel, 1998). Furthermore, it is known that the phoneme discrimination ability in these children can be increased by training with more salient versions of the rapidly changing elements in the acoustic waveform of speech (Merzenich *et al.,* 1996; Tallal *et al.,* 1996).

It is a well-evidenced theory that an impairment of temporal resolution may be one reason for underlying language deficits at the phonological level (Bishop, 1997; Tallal and Newcombe, 1978). As mentioned above, a fine-grained auditory resolution is required to decode spoken language. It can be assumed that a general timing deficit in patients with aphasia caused the specific problem in detecting temporal characteristics of the speech signal (Von Steinbuchel *et al.,* 1999). Analogously, language learning-impaired

children display a disturbance in temporal resolution and phoneme discrimination (Breier *et al.,* 2001).

The coexistence of impairments at a phonological-language level and reduced discriminating abilities for non-speech sounds with rapid temporal characteristics in these patients gives evidence for the hypotheses that the underlying deficits might be associated with inappropriate functioning of a universal system for auditory temporal resolution. This suggestion is strongly linked to the general consideration of the uniqueness or generality of the mechanism underlying rapid temporal processing of speech and non-speech sounds.

1.3 Behavioral research on the perception of temporal cues

Research on cerebral asymmetries traditionally reports a right ear, respectively left hemisphere advantage for the discrimination of speech stimuli. In parallel, a right ear advantage has been reported for the perception of non-speech stimulation including perception of temporal order (Mills and Rollman, 1980), duration discrimination (Mills and Rollman, 1979), discrimination offsets of tones (Emmerich *et al.,* 1981) and gap detection (Brown and Nicholls, 1997).

1.3.1 The Perception of Voice-onset-time (*VOT*)

Voice onset time is one of several temporal cues that is utilized in our perception and identification of stop consonants. This cue reflects the temporal relationship between laryngeal pulsing (e.g., vocal chord vibration) and consonant release (e.g. the separation of lips to release a burst of air from the vocal tract during the production of bilabial stop consonants such as /b, p/). Behavioral investigations have indicated that when the VOT is systematically manipulated, adult listeners are able to discriminate changes in VOT only to the extent that they can assign unique labels to these sounds (Liberman *et al.,* 1967). Thus, continuous changes in VOT are generally perceived categorically. This effect has been demonstrated in human adults (Liberman *et al.,* 1958; Lisker and Abramson, 1964; Lisker and Abramson,

1970) as well as in children and young infants (Eimas *et al.,* 1971; Treub and Rabinovitch, 1972). Combinations of an initial plosive and followed by a vowel are the most frequent syllable structures. These stop consonant vowel syllables occurs in almost all languages of the world and are the first linguistic unit to develop in ontogenesis (Kuhl, 2004; Ladefoged and Maddieson, 1997; Oller, 2000).

1.3.2 The Perception of Noise-onset-time (*NOT*) in "Gap-detection"

The most common task used to behaviorally tap auditory temporal resolution is gap detection. The listener is usually presented with two streams of sound, one of which has a brief silent period ("gap") at its temporal midpoint. The task of the listener is to identify this signal and some form of adaptive tracking procedure (Levitt, 1971) is used to determine the shortest detectable gap ("gap threshold"). A variation of this paradigm, which has been termed "between-channel" gap detection, requires the listener to detect a gap bounded by markers (leading and trailing elements) that differ in frequency content, ear stimulation, or location in free-field space. The gap to be detected is delimited by markers differing in spectral or spatial qualities. This requires a relative timing operation to be performed on the activity between different perceptual channels. The gap thresholds are longer in this paradigm and they increase as the leading marker is shortened (Phillips *et al.,* 1997; Phillips *et al.,* 1998).

It is assumed that the processing of this task requires the same perceptual mechanisms involved in stop consonant discrimination. In the case of VOT processing, the task is to recognize the relative timing of the high frequency consonantal burst and the low frequency vowel (Phillips and Smith, 2004). It is shown that for short leading elements (5ms), the gap-threshold is around 30ms (Phillips *et al.,* 1997), which is close to the VOT of the perceptual boundary between some voiced and voiceless consonants (Eggermont, 1995; Kuhl and Miller, 1978). Furthermore, a right-ear-advantage analogous to the well known right-ear-advantage in phoneme discrimination, could be detected when presenting the gap-stimuli to

different ears. (Vroom *et al.,* 1977). Additionally, a recent study by Sulakha and co-workers showed a left hemisphere advantage in auditory gap-detection paradigm when gaps of silence are presented laterally within white (broad-band) noise (Sulakhe *et al.,* 2003). According to the term voice-onset-time (VOT), defined as the time between release of closure and start of voicing in CV-syllables, here we use the term "Noise-onset-time" for the time (silent gap) between the trailing and leading noise elements of the non-speech sounds.

In terms of rapid temporal information processing a relationship between speech and non-speech perception has also been demonstrated using a tone onset time series (TOT). The TOT series is a well established non-speech analogue of a VOT series (Pisoni, 1977). TOT stimuli consist of two tones that are either presented simultaneously or with various onset asynchronies that mimic the voice onset times. Participants are asked to assign the stimuli to one of two categories, either "together" or "apart". Studies using this paradigm in adults showed evidence for a category boundary with a critical range around 30ms (Pisoni, 1977). Children with reading disabilities exhibited reduced categorical perception in VOT as well as TOT discrimination (Breier *et al.,* 2001).

Taken together these behavioral studies provide evidence for a strong relation between language-specific phonological and non-specific rapid temporal information processing. As a hemispheric preponderance seems to exist for processing phonological stimuli, it is plausible to assume that the left hemisphere's preference for rapid temporal information actually accounts for this result.

1.4 Evidences from neuroimaging

Using positron-emission-tomography (PET) and functional magnetic resonance imaging (fMRI) numerous studies have delineated the various specializations of the human auditory system (Belin *et al.,* 1999; Belin *et al.,* 2000; Binder *et al.,* 2000; Griffiths *et al.,* 1998; Griffiths and Warren, 2002;

Jancke *et al.,* 2001; Jancke *et al.,* 2002; Jancke *et al.,* 2003; Jancke and Shah, 2002; Meyer *et al.,* 2002; Meyer *et al.,* 2003; Meyer *et al.,* 2005; Warren *et al.,* 2006; Zatorre *et al.,* 2002a; Zatorre *et al.,* 2002b; Zatorre and Belin, 2001). The imaging research has typically found specific regions that show speech sensitivity, especially the left posterior portion of the superior temporal gyrus (STG) and sulcus (STS) in both hemispheres, which were activated while listeners processed acoustic speech compared to non-speech stimulation. One interpretation of these results is that this region is specifically involved in processing the phonetic form of speech and therefore mediates rapid temporal processing. Further results support this view of functional specialization of the superior temporal cortex for rapid temporal processing.

Further investigations into the role of temporal processing in language lateralization used passive auditory stimulation with nonverbal sounds containing either rapid or slow frequency transitions. A left biased hemispheric asymmetry could be observed in response to rapid frequency transitions due to a reduced response of the right auditory cortex. The authors consider their finding to be evidence for a basic component of hemispheric lateralization that depends on temporal properties of auditory input, rather than depending on additional verbal or phonetic operations. Therefore this data support the view that temporal processing may contribute to language lateralization at a prelinguistic stage (Belin *et al.,* 1998).

A further investigation used speech and commensurating with non-speech sounds to study the processing of temporally based acoustic features via records of intracerebral evoked potentials. The results showed a left lateralized processing of rapidly changing acoustic elements in speech and non-speech sounds in the auditory cortex (Liegeois-Chauvel *et al.,* 1999). A recent MEG study uncovered a left-lateralized activation of the posterior portion of the superior temporal gyrus for the processing of VOT-varied stimuli and no hemispheric asymmetries in the response to TOT modulated stimuli (Papanicolaou *et al.,* 2003). A further brain imaging study examining the anatomical basis of rapid temporal discrimination in the auditory system

failed to point to any hemispheric specialization In this work the bilateral posterior portion of the superior temporal sulcus exhibit greater activation for speech and non-speech stimuli that involved rapid temporal discrimination than for speech and non-speech stimuli involving spectral discrimination (Joanisse and Gati, 2003).

The findings sketched above clearly indicate that left cortical areas are essential for phonological analysis of speech sounds. Still unclear however, is the extent to which these regions are also critically involved in lower-level analysis of speech stimuli. Furthermore, it is as yet not known whether the role of these areas is restricted to the analysis of speech stimuli or whether it also extends to the analysis of non-speech sounds with rapidly changing temporal information.

1.5 Methods: Functional magnetic resonance imaging (fMRI) and electroencephalography (EEG)

In this chapter, the two brain imaging methods used in the work presented will be described.

Functional magnetic resonance imaging (fMRI) provides images of the brain function based on information on metabolism and blood perfusion. Therefore, this method measures the neuronal activity indirectly with very high spatial resolution in the range of millimetres, but with low temporal resolution in the range of seconds. The time course of the fMRI hemodynamic response is roughly a low pass filtered version of the electric neuronal activity. (Logothetis *et al.,* 2001).

In contrast, electroencephalogram (EEG) measures the neuronal electrical activity directly with an excellent time resolution with a range of milliseconds, but low spatial resolution with a range of centimetres.

The present work combines both research methods with the described complementary advantages to achieve a good temporal and a good spatial resolution.

1.5.1 Functional magnetic resonance imaging (fMRI)

Most of the studies in this work used functional magnetic resonance imaging (fMRI) to assess human brain functions. This method examines hemodynamic changes in response to specific sensory, cognitive or emotional states to map the spatiotemporal distribution of neuronal activity. The main advantages of fMRI as a technique used to image brain activity include that the procedure does not require injections of radioactive isotopes and that the spatial resolution of the functional image is generally very high (usually about 1.5 x 1.5 mm, although resolutions of less than 1 mm are possible).

The subsequent sections discuss the basic principles of MRI scanning and commonly used statistical methods for fMRI data analysis.

1.5.1.1 Basic principles of MRI scanning

Magnetic resonance imaging (MRI) emerged from a physical phenomenon known as Nuclear Magnetic Resonance (NMR), which deals with the magnetic properties of the nucleus of atoms. NMR is a phenomenon which occurs when the nuclei of certain atoms are immersed in a static magnetic field and exposed to a second oscillating magnetic field.

The basic principles of the MRI account for i.) the magnetic spin and the larmor frequency, ii) the signal detection, iii) the relaxation times, and iv) the BOLD contrast, which will be discussed in turn.

Spin and Larmor frequency

A basic characteristic of all elements is their *spin*. Spin is a fundamental property of nature like electrical charge or mass. Spin comes in multiples of 1/2 and can be positive or negative. Protons, electrons and neutrons possess spin, and individual unpaired electrons, protons, and neutrons each possesses a spin of 1/2. Only nuclei with an uneven number of protons result in a detectable spin. Therefore, the described technique deals with the nucleus of a hydrogen atom, which consists of a single proton and furthermore, constitutes the biggest source of protons in the body.

Without an external magnetic field the magnetic moments of the hydrogen nuclei are randomly distributed; the net magnetization is zero (**M=0**). Under the impact of a strong external magnetic field ($\mathbf{B_0}$), the protons align with the external field in parallel or antiparallel fashion. They obey the Boltzmann statistics and therefore a slight majority aligns parallel with the magnetic field, because parallel alignment with the external field leads to a lower energy state and is therefore preferred. The summation of the individual magnetic moments leads to a net longitudinal magnetization ($\mathbf{M_z}$) along the external magnetic field.

In an external magnetic field ($\mathbf{B_0}$) protons precess around their vertical axis with a particular resonance frequency, the so-called larmor frequency ($\boldsymbol{\omega_o}$,) (cf. Figure 1). This larmor frequency is directly proportional to the strength of the magnetic field [$\mathbf{B_0}$ in Tesla] and to a constant γ, which is characteristic for each proton. The larmor frequency is:

$$\omega_0 = \gamma B_0$$

With **γ** being a constant unique to every atom (gyromagnetic ratio) and $\mathbf{B_0}$ being the main magnetic field, measured in Tesla.

All protons in this magnetic field have an additive magnetic potential and therefore produce a net longitudinal magnetization ($\mathbf{M_z}$) in the direction of the magnetic field $\mathbf{B_0}$**.** Thus, once $\mathbf{M_z}$ is built up, the system is in equilibrium.

Signal detection

At equilibrium, the net magnetization vector lies along the direction of the applied external magnetic field $\mathbf{B_o}$ and is called the equilibrium magnetization. In this configuration, the Z component of magnetization $\mathbf{M_Z}$ is referred to as the longitudinal magnetization. A transverse ($\mathbf{M_X}$ or $\mathbf{M_Y}$) magnetization does not exist. Figure 1 illustrates the orientation of the magnetization components.

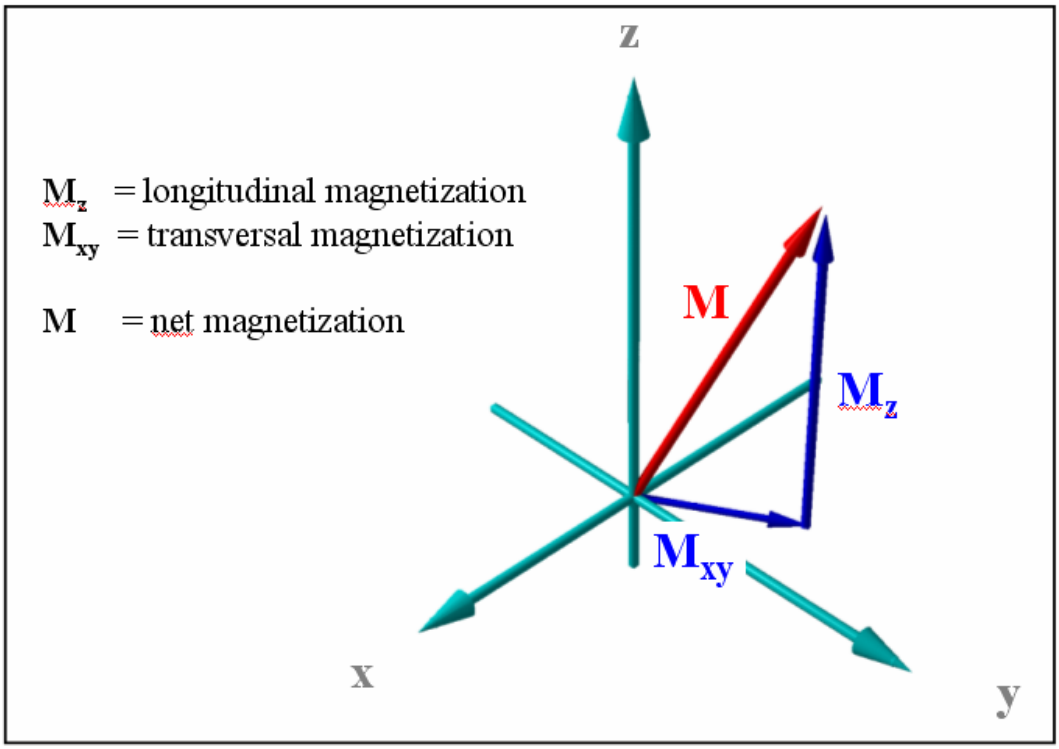

Figure 1: Orientation of the coordinate system and the magnetization components longitudinal magnetization ($\mathbf{M_z}$), transversal magnetization ($\mathbf{M_{xy}}$), and resulting net magnetization ($\mathbf{M}$).

It is now possible to change the net magnetization by applying an electromagnetic radio frequency (RF) pulse with the larmor frequency of the specific protons. This leads to a saturation of the spin system and the longitudinal magnetization spirals down towards the XY plane and, if the applied energy is sufficient, $\mathbf{M_Z}$ becomes zero. Therefore, the RF pulse results in a decrease of the net magnetization ($\mathbf{M_z}$) perpendicular to the longitudinal field and in the formation of the transversal magnetization ($\mathbf{M_{xy}}$). $\mathbf{M_{xy}}$ is constantly rotating in the xy-plane around the z-axis with the frequency ($\boldsymbol{\omega_o}$). This rotation of the transversal magnetization leads to an electromagnetic induction on a receiver coil and can be measured.

The induced transversal magnetization ($\mathbf{M_{xy}}$) only has a short durability. The decrease of the $\mathbf{M_{xy}}$ is caused by "spin-lattice-interactions" and "spin-spin-interactions". The "spin-lattice-interactions" produce a decay of the energy back to the equilibrium whereby the absorbed RF energy is

retransmitted at the larmor frequency, producing the NMR signal. This leads to a recovery of the longitudinal magnetization **M_Z** (*longitudinal relaxation*). The time constant which describes how **M_Z** returns to its equilibrium value is called the spin lattice relaxation time (**T1**). **T1** is tissue specific and depends on the external magnetic field (**B_0**). **T1** is defined as the time after 63% of the longitudinal magnetization is recovered. Figure 2 shows a prototypical T1- time course.

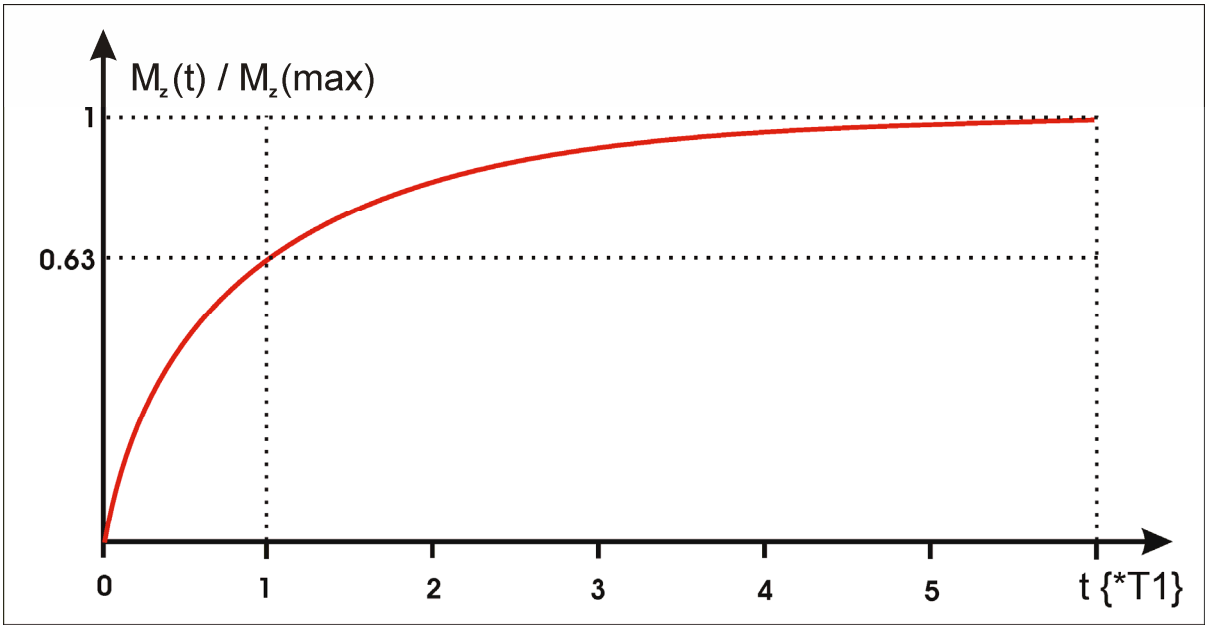

Figure 2: Longitudinal relaxation (T1). Graph shows the recovery of the longitudinal magnetization (M) as a function of time (t in units of T1). Dotted lines indicate time points, when 63% of the longitudinal magnetization is recovered (T1) and when longitudinal magnetization is completely recovered.

During the rotation of **M_{xy}** around the *z*-axis, the transversal magnetization starts to dephase because each of the individual spins that make up this magnetization experiences a slightly different magnetic field. The process of dephasing of the excited protons or spins is called the "spin-spin-interaction". Along with this loss of synchronicity the transversal magnetization converges to zero (transversal relaxation). In other words, the temporary and random interaction between spins cause a cumulative loss in phase and results in an overall loss of transversal magnetization, also called transverse relaxation (also spin or T2-relaxation; T2-decay). The time constant which describes the return to equilibrium of the transverse

magnetization (**M_{XY}**) is called the spin-spin relaxation time (**T2)**. **T2** is defined as the time after the transversal magnetization decreased on 37% of its initial value. Figure 3 shows a prototypical T2- time course. Unlike **T1** values, **T2** is not related to field strength.

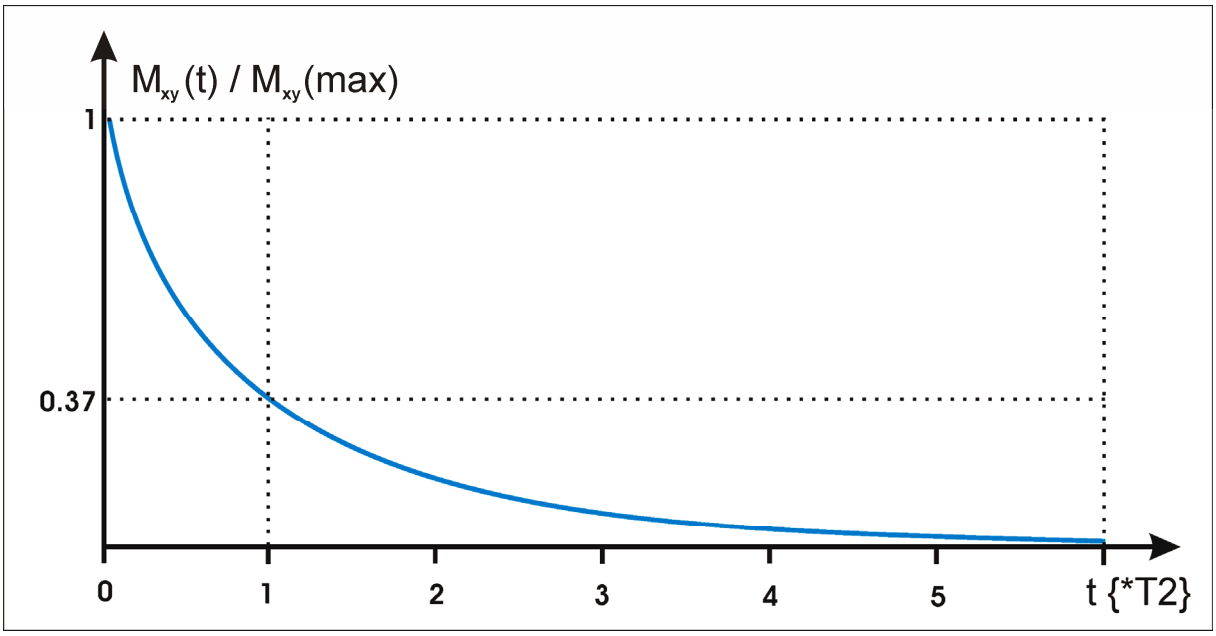

Figure 3: Transversal relaxation (**T2**). Graph shows the dephasing of the transversal magnetization (**M_{xy}**) as a function of time (**t** in units of **T2**). Dotted lines indicate the two time points when the transversal magnetization is decreased on 37% of its initial value (**T2**) and when the transversal magnetization is completely decreased.

Pure **T2** decay is a function of random interactions between spins. The assumption is that the main external field **B_0** is absolutely homogeneous, but many factors can lead to imperfections in the homogeneity of a magnetic field. The sum total of all these random and systematic effects lead to a much faster transversal relaxation, which is defined by the time constant **T2***.

The BOLD contrast

The MR signal that corresponds to the neural activity is called the blood oxygen level-dependent (BOLD) signal change or BOLD response. Functional MRI is based on the increase in blood flow to the local vasculature that accompanies neural activity in the brain. This results in a

corresponding local reduction in deoxyhemoglobin because the increase in blood flow occurs without a similar increase in oxygen extraction. Since deoxyhemoglobin is paramagnetic, it alters the T2* weighted magnetic resonance image signal (Ogawa, et al, 1990a and b, 1992, 1993). Thus, deoxyhemoglobin is referred to as an endogenous contrast enhancing agent and serves as the source of the signal for fMRI.

The time course of the hemodynamic response is of particular interest (cf. Figure 4). After an initial decrease of oxyhemoglobin due to the consumption of oxygen (initial dip) there is an increase in blood flow and therefore an *increase* in *ox*yhemoglobin in relation to deoxyhemoglobin is observed. The peak of this increase is reached 5-8 seconds after the stimulus onset that is assumed to cause the neural activity and returns to baseline after 10-16 seconds. This paradoxical increase of oxyhemoglobin due to neural activity has been described by Fox and Raichle (1986). It is that change in the ratio of deoxyhemoglobin to oxyhemoglobin that is measured by fMRI techniques**.**

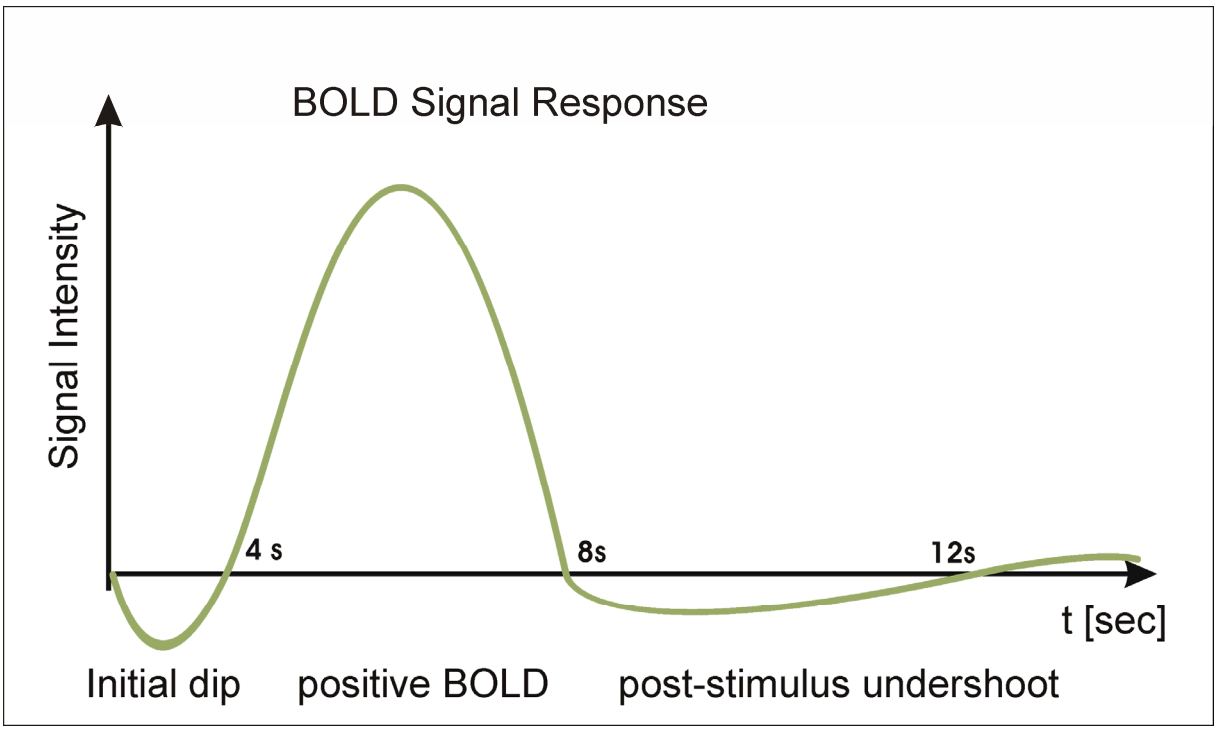

Figure 4: The time course of the hemodynamic response.

The blood hemoglobin has different magnetic characteristics depending on its state of oxygenation. While oxygenated hemoglobin has only a small magnetic moment and is diamagnetic, deoxygenated hemoglobin is paramagnetic. That is, deoxygenated hemoglobin behaves like a small magnet due to the iron in its molecules. That slight difference in the relation between the magnetic characteristics of oxyhemoglobin and deoxyhemoglobin can be detected in the MR signal. In particular, the decrease of deoxyhemoglobin leads to a decrease of the slowing down of the transversal dephasing and therefore results in an increasing T2* relaxation time and as a result in the enhancement of the MR- signal.

1.5.1.2 Methods of analysis

Preprocessing in fMRI

The purpose of preprocessing in fMRI is to correct for non-task related variability in experimental data. These approaches are usually performed regardless of the experimental design and therefore are called preprocessing. The preprocessing steps seek to remove rather than model data variability. One important term in fMRI analysis language is the signal-to-noise-ratio (SNR), which is the quotient between task-related variability and non-task-related variability. The goal of the preprocessing is to remove as much non-task-related variability as possible in order to obtain a high SNR value. The sensitivity of an fMRI analysis is determined by the amount of residual noise (non-task related variability) in the image series. The following sections describe three basic steps of preprocessing (although the number of steps depends on the applied experimental design).

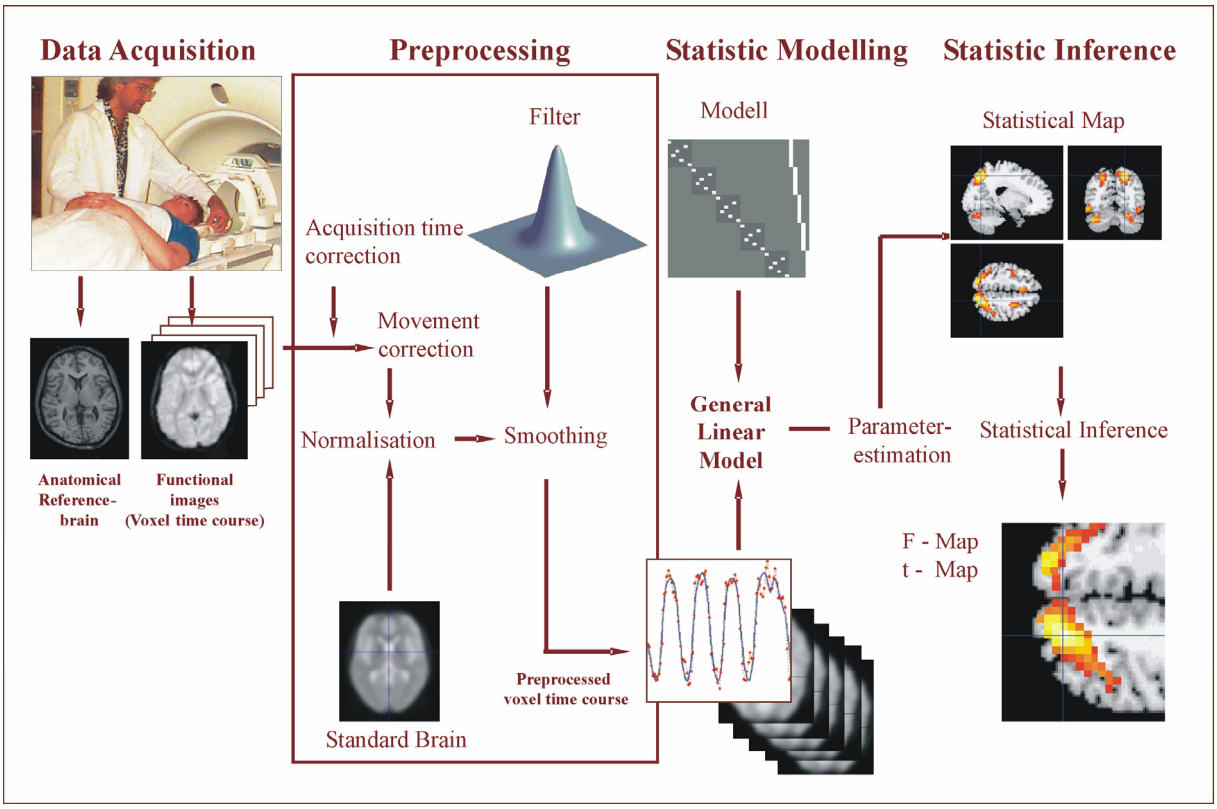

Figure 5: Analysis steps for fMRI data (modified from Franckowiak et al. 1997)

Motion correction

Motion correction corrects the data for movements of the head; it controls for motion-related variance. Small movements of approximately 1mm occur (Turner et al., 1997; Franckowiak et al., 1997) even if the subject's head is fixated and the subject explicitly tries not to move. In addition, there are movements related to physiological factors, such as cardiac and pulmonary parameters. These movements can e.g. alter the "spin history" as a result to the change of position relative to the gradients of the scanner. Moreover, it is often impossible to eliminate the voxels in the activation pattern that are activated by an artifact that has arisen as a result of motion. The functional volumes are aligned in space by rigid-body transformations. The software program SPM employs a least-squares approach and a 6 parameter (rigid body) spatial translation (Friston et al. 1996). In this approach, the first image in a time-series serves as a reference

scan against which all subsequent scans are aligned. Motion correction has several limitations, including artifact-related limitations, which can result in a loss of data around the edges of the image volume. Distortions in fMRI images can also pose a problem because they may be dependent upon position in field, rather than position in the head.

Normalization

This preprocessing step spatially (stereotactically) normalizes images into a standardized space, which is defined by some ideal model or template image. The software program SPM conforms to the Montreal Neurological Institute (MNI) space (a combination of 152 MRI scans on normal, right-handed subjects) and approximates that of the space described in Talaraich and Tournoux's atlas (1988). Generally, these algorithms work by minimizing the sum of squares differences between the acquired images and the template (for automated algorithms see e.g. Ashburner & Friston, 1997, 1999). One major advantage of the normalization is that it allows the generalization of results to a larger population, thereby enabling averaging across subjects and improving comparisons with other studies. Normalization also provides a coordinate space to report the results. One disadvantage of normalization can be cited in its potential to reduce the spatial resolution resulting from differences in interindividual brain organization and thus may reduce the activation strength by averaging subjects.

Spatial Smoothing

Spatial smoothing convolves images with an isotropic Gaussian kernel (defined by its full width at half maximum (FWHM)) to potentially increase SNR, to compensate for residual between-subject variability after normalization and to approximate a random field for statistical purposes. Issues such as reduced spatial resolution, which relate to smoothing, are discussed in detail in Friston et al., 2000.

Statistical analysis of fMRI data

Several methods based on the General Linear Model (GLM) can be employed to assess brain activity in fMRI studies. The basic concept of the GLM is that it treats the data as a linear combination of model functions plus noise. Typically, these methods result in a statistical parametric map. In order to obtain these parametric maps, each of these methods entails comparable successive steps.

Firstly, the response (signal) will be modeled at each voxel in the brain by a GLM. Subsequently, a hypothesis regarding the model's parameters will be tested. The observed statistical map will then be represented at a given level (threshold) according to the point of distribution of the statistics (uncorrected level) or the field distribution of the statistics (corrected levels for local maxima). A correction to the significance of the t-statistics is suggested, which account for the multiple comparisons in the image. Afterwards, the statistical parametric map can be visualized showing e.g. the t-value for each single voxel with a corresponding color on an image of the brain. Overall, these methods are voxel-by-voxel hypothesis testing approaches, which reliably identify regions showing a significant effect of interest.

The GLM model used can refer to a single subject, a group of subjects or multiple groups of subjects. For a simple t-test, the GLM is also utilized in order to account for covariates as well as temporal autocorrelation. Multi-subject fMRI experiments can also be performed using a GLM framework with different forms depending on the approach taken, e.g. fixed or random subject analysis. A fixed-effects model allows interference in the analyzed subject sample. A random-effects model accounts for inter-subject variance in analyses and permits interference for the population from which subjects are drawn, which is essential for group comparisons.

1.5.2 Electroencephalography (EEG)

Electroencephalography is the measurement of the electrical activity of the brain as recorded from electrodes placed on the scalp. The resulting traces are known as an electroencephalogram (EEG) and represent an electrical signal (postsynaptic potentials) from a large number of neurons. In 1929 Hans Berger published the first paper about the human electroencephalogram, describing the recording of electric brain signals measured on the human scalp (Berger, 1929). In comparison to other research methods in neuroscience

EEG has several advantages as an instrument for exploring human brain activity. EEG has a very high temporal resolution down to sub-millisecond. Furthermore, in contrast to other methods for exploring cerebral functions that rely on blood flow or metabolism and are therefore decoupled from the brain's electrical activity, EEG is a method to measure it directly.

The following section will briefly discuss the topics of EEG recording, the basic physiological principles of the EEG signal, the concept of event related potentials, and methods to analyze EEG data.

1.5.2.1 Recording of human scalp EEG

To compare results over different studies, electrode locations must be standardized. For that reason, in the recording of conventional scalp EEG electrodes were placed on the scalp using standardized placement systems such as the 10/20-system (Jasper, 1958). Here electrodes are placed at fixed distances according to ten or twenty percent distance between distinctive landmarks on a subject's head (nasion, inion, and left and right preauricular points). In the EEG-study of the present work, a modification of this standard system, the 10/10 system with 32 electrodes, was used (cf. Figure 6). Here the distances between the electrodes average ten percent the distance between the particular landmarks.

The EEG signal consists of the scalp potential differences between active electrodes and the reference electrode (either a physical reference

electrode or any recomputed reference). Generally, electric potentials are only defined with respect to a reference. Therefore, for each EEG recording, a reference electrode has to be selected in advance. This electrode should be affected by global voltage changes in the same manner as all the other electrodes, such that brain unspecific activity is subtracted out by the referencing. Also, the reference should not additionally record unintended signals, like heart activity. Commonly, a reference on the head but at some distance from the other recording electrodes is chosen; the ear lobes, the nose, or the mastoids, for example. In the present study, we used a computed "average reference"; a subtraction of the average over all electrodes from each electrode for each time point. This distributes the "responsibility" over all electrodes, rather than assigning it only to one of them.

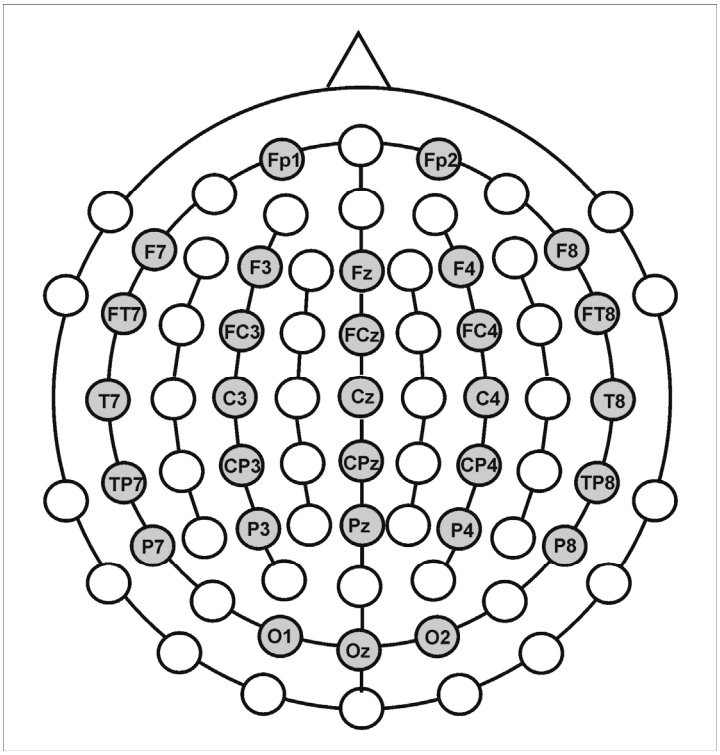

Figure 6: 10/10 standard system. Locations of 30 electrodes following the 10/10 system

1.5.2.2 The electrophysiological basis of EEG

The electrical activity of active nerve cells in the brain produces currents spreading through the head. These currents also reach the scalp surface and the resulting voltage differences on the scalp can be recorded as

the electroencephalogram (EEG). EEG reflects brain electrical activity with millisecond temporal resolution, is a direct correlate of brain and can be obtained non-invasively. If an electrical signal is transmitted along an axon or dendrite, electrical charges are separated along very short distances on the corresponding cell membranes. Separated charges produce positive polarity at the side where the current is leaving and negative polarity where the current is returning. These effects are called "primary currents" since they reflect the activity of the corresponding neurons directly. The primary currents are embedded in an environment of brain tissue and liquor. Therefore, a current is induced which flows through these media, as well as the skull and scalp, called a "secondary" current. Secondary currents also reach the scalp surface and cause voltage differences at the scalp that can be picked up by EEG electrodes.

Contributors to the measurable EEG signals are action potentials along the axons connecting neurons, currents through the synaptic clefts connecting axons with neurons and dendrites, and currents along dendrites from synapses to the soma of neurons. However, the current flow along post-synaptical apical dendrites in cortical pyramidal cells of the cortex is assumed to contribute strongest to the measurable EEG signals. Yet, current in a single neuron is small in magnitude and a considerable distance from the scalp surface, so one dendrite would be far too weak to produce a measurable signal; instead thousands of synchronously activated neurons are required. Consequently, EEG is only sensitive to coherent simultaneous activity of a large number of neurons.

Two classes of activation can be distinguished. The "spontaneous" neural activity reflects continuous brain activity. In contrast, activity elicited by internal or external stimulation, the "event-related potentials" (ERP), are constituted of averaged transient electrical potentials that are time-locked to the repeated presentation of discrete stimuli.

1.5.2.3 Event-related potentials (ERPs)

Event-related potentials (ERPs) are brain responses time-locked to an "event" (internal or external stimulus). In the present work, auditory related potentials (ARPs) were used. Figure 7 illustrates a prototypical ERP in response to an auditory event. AEPs are electrical voltage potentials originating from the brain recorded from the scalp in response to an auditory stimulus. Within the ARPs, different responses can be distinguished. The “Auditory Brainstem Response” (ABR) occurs around 1.5-15 ms after stimulus presentation. The ABR originates in the cranial nerve and brainstem auditory structures. The “Middle Latency Response” (MLR) occurs 25-50 ms after stimulus presentation and originates in upper brainstem and auditory cortex. The “Slow” cortical auditory ERPs, including the P1-N1-P2 complex occur 50-200 ms after stimulus presentation and originate in the auditory cortex. The N1 is a large negative wave that occurs 80-100 ms after stimulus presentation and originates primarily in the bilateral auditory cortex. The “Late” cortical auditory ERPs, including the “Mismatch Negativity” (MMN) start around the time of N1 *and* later. The MMN is a response reflecting detection by the brain of a *change* in the stimulus. Other "late" ERPs, include N2b and P3, which are cortical ERPs but not specifically related to auditory structures. These responses are characteristic "components" in their time course of the ARP.

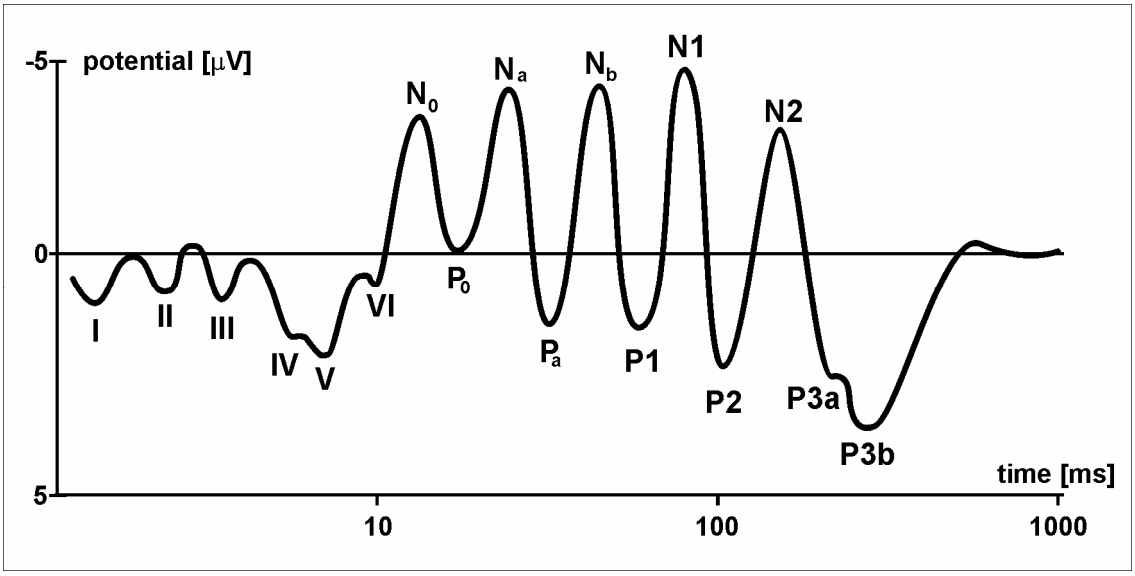

Figure 7: Prototypical auditory related potential ARP. Figures shows an event-related potential (ERP) in response to an auditory event as a function of time.

1.6 Reference List

Amaro, E., Williams, S.C., Shergill, S.S., Fu, C.H., MacSweeney, M., Picchioni, M.M., Brammer, M.J., McGuire, P.K., 2002. Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. J.Magn Reson.Imaging 16, 497-510.

Ashburner, J. & Friston, K. (1997). Multimodal image coregistration and partitioning– a unified framework. Neuroimage, 6(3): 209-217.

Ashburner, J. & Friston, K. (1999). Nonlinear spatial normalization using basic functions. Hum. Brain Mapp., 7(4): 254-266.

Belin, P., Zatorre, R.J., Hoge, R., Evans, A.C., Pike, B., 1999. Event-related fMRI of the auditory cortex. Neuroimage. 10, 417-429.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309-312.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.C., Samson, Y., 1998. Lateralization of speech and auditory temporal processing. J.Cogn Neurosci. 10, 536-540.

Berger, H. (1929). Über das Elektronencephalogramm des Menschen. *Archiv für Psychiatrieund Nervenkrankenheit, 87,* 527-570.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb.Cortex 10, 512-528.

Bishop, D.V., 1997. Language impairment. Listening out for subtle deficits. Nature 387, 129-130.

Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat.Neurosci. 8, 389-395.

Breier, J.I., Gray, L., Fletcher, J.M., Diehl, R.L., Klaas, P., Foorman, B.R., Molis, M.R., 2001. Perception of voice and tone onset time continua in children with dyslexia with and without attention deficit/hyperactivity disorder. J.Exp.Child Psychol. 80, 245-270.

Brown, S., Nicholls, M.E., 1997. Hemispheric asymmetries for the temporal resolution of brief auditory stimuli. Percept.Psychophys. 59, 442-447.

Bryden, P., 1982. Laterality: Functional Asymmetry in the Intact Brain. Academic Press, Ney York.

Efron, R., 1963. Temporal Perception, Aphasia and D'ej'a vu. Brain 86, 403-424.

Eggermont, J.J., 1995. Neural correlates of gap detection and auditory fusion in cat auditory cortex. Neuroreport 6, 1645-1648.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., Vigorito, J., 1971. Speech perception in infants. Science 171, 303-306.

Emmerich, D.S., Pitchford, L.J., Joyce, C.M., Koppell, S.M., 1981. Laterality effects in response to offsets of tonal stimuli. Neuropsychologia 19, 227-234.

Fiez, J.A., Raichle, M.E., Miezin, F.M., Petersen, S.E., Tallal, P., Katz, W.F., 1995. Pet Studies of Auditory and Phonological Processing - Effects of Stimulus Characteristics and Task Demands. Journal of Cognitive Neuroscience 7, 357-375.

Franckowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J. & Mazziotta, J.C. (1997). Human brain function. San Diego: Academic Press.

Friston, K.J., Williams, S., Howard, R., Franckowiak, R.S. & Turner, R. (1996). Movement-related effects in fMRI time-series. Magn. Reson. Med., 35 (3): 346-355.

Friston, K.J., Josephs, O., Zarahn, E., Holmes, A.P., Rouquette, S. & Poline, J. (2000). To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. Neuroimage, 12(2): 196-208.

Griffiths, T.D., Buchel, C., Frackowiak, R.S., Patterson, R.D., 1998. Analysis of temporal structure in sound by the human brain. Nat.Neurosci. 1, 422-427.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348-353.

Haller, S., Bartsch, A.J., Radue, E.W., Klarhofer, M., Seifritz, E., Scheffler, K., 2005. Effect of fMRI acoustic noise on non-auditory working memory task: comparison between continuous and pulsed sound emitting EPI. MAGMA. 18, 263-271.

Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci. 4, 131-138.

Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67-99.

Jancke, L., Buchanan, T.W., Lutz, K., Shah, N.J., 2001. Focused and nonfocused attention in verbal and emotional dichotic listening: an FMRI study. Brain Lang 78, 349-363.

Jancke, L., Shah, N.J., 2002. Does dichotic listening probe temporal lobe functions? Neurology 58, 736-743.

Jancke, L., Specht, K., Shah, J.N., Hugdahl, K., 2003. Focused attention in a simple dichotic listening task: an fMRI experiment. Brain Res.Cogn Brain Res. 16, 257-266.

Jancke, L., Wustenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. Neuroimage. 15, 733-746.

Jasper, H. (1958). The ten twenty electrode system of the International Federation.*Electroencephalography and Clinical Neurophysiology, 10,* 371-375.

Joanisse, M.F., Gati, J.S., 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. Neuroimage. 19, 64-79.

Kimura, D., 1961. Some effects of temporal-lobe damage on auditory perception. Can.J.Psychol. 15, 156-165.

Kimura, D., 1964. Left-right differences in the perception ofmelodies. Quarterly Journal of Experimental Psychology 14, 355-358.

Kimura, D., Folb, S., 1968. Neural processing of backwards-speech sounds. Science 161, 395-396.

King, F., Kimura, D., 1972. Left-ear superiority in dichotic perception of vocal nonverbal sounds. Canadian Journal of Psychology 26, 111-116.

Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. Nat.Rev.Neurosci. 5, 831-843.

Kuhl, P.K., Miller, J.D., 1978. Speech perception by the chinchilla: identification function for synthetic VOT stimuli. J.Acoust.Soc.Am. 63, 905-917.

Ladefoged, P., Maddieson, I., 1997. The Sounds of the World's Languages. Blackwell, Oxford.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J.Acoust.Soc.Am. 49, Suppl.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. Psychol.Rev. 74, 431-461.

Liberman, A.M., Delattre, P.C., Cooper, F.S., 1958. Some cues for the distinction between voiced and voiceless stops in initial position. Language and Speech 1, 153-167.

Lichtheim L., 1885. On aphasia. Brain 7, 433-484.

Liegeois-Chauvel, C., de Graaf, J.B., Laguitton, V., Chauvel, P., 1999. Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb.Cortex 9, 484-496.

Lisker, L., Abramson, A.S., 1964. Across language study of voicing in initial stops: Acoustical measurements. Word 20, 384-411.

Lisker, L., Abramson, A.S., 1970. The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences. Academia, Prague, pp. 563-567.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature, 412, 150-157.

Mates, J., Steinbuchel, N., Wittmann, M., Treutwein, B., 2001. A system for the assessment and training of temporal-order discrimination. Computer Methods and Programs in Biomedizin 64, 125-131.

Merzenich, M.M., Jenkins, W.M., Johnston, P., Schreiner, C., Miller, S.L., Tallal, P., 1996. Temporal processing deficits of language-learning impaired children ameliorated by training. Science 271, 77-81.

Meyer, M., Alter, K., Friederici, A., 2003. Functional MR imaging exposes differential brain responses to syntax and prosody during auditory sentence comprehension. Journal of Neurolinguistics 16, 277-300.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., 2002. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum.Brain Mapp. 17, 73-88.

Meyer, M., Zysset, S., von Cramon, D.Y., Alter, K., 2005. Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. Brain Res.Cogn Brain Res. 24, 291-306.

Mills, L., Rollman, G.B., 1979. Left hemisphere selectivity for processing duration in normal subjects. Brain Lang 7, 320-335.

Mills, L., Rollman, G.B., 1980. Hemispheric asymmetry for auditory perception of temporal order. Neuropsychologia 18, 41-48.

Moelker, A., Pattynama, P.M., 2003. Acoustic noise concerns in functional magnetic resonance imaging. Hum.Brain Mapp. 20, 123-141.

Nunez, P. L. (1981). *Electrical fields of the brain*. New York: Oxford University Press.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Science of the USA, 87, 9868-9872.

Oller, D.K., 2000. The Emergence of the Speech Capacity. Erlbaum, Mahwah.

Papanicolaou, A.C., Castillo, E., Breier, J.I., Davis, R.N., Simos, P.G., Diehl, R.L., 2003. Differential brain activation patterns during perception of voice and tone onset time series: a MEG study. Neuroimage. 18, 448-459.

Phillips, D.P., Hall, S.E., Harrington, I.A., Taylor, T.L., 1998. "Central" auditory gap detection: a spatial case. J.Acoust.Soc.Am. 103, 2064-2068.

Phillips, D.P., Smith, J.C., 2004. Correlations among within-channel and between-channel auditory gap-detection thresholds in normal listeners. Perception 33, 371-378.

Phillips, D.P., Taylor, T.L., Hall, S.E., Carr, M.M., Mossop, J.E., 1997. Detection of silent intervals between noises activating different perceptual channels: some properties of "central" auditory gap detection. J.Acoust.Soc.Am. 101, 3694-3705.

Pisoni, D.B., 1977. Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. J.Acoust.Soc.Am. 61, 1352-1361.

Poeppel, D., 2001. Pure word deafness and the bilateral processing of the speech code. Cognitive Science 25, 679-693.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication 41, 245-255.

Robin, D.A., Tranel, D., Damasio, H., 1990. Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. Brain Lang 39, 539-555.

Schwartz, J., Tallal, P., 1980. Rate of acoustic change may underlie hemispheric specialization for speech perception. Science 207, 1380-1381.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 Pt 12, 2400-2406.

Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100-107.

Steinbuchel, N.v., Poppel, E., 1993. Domains of rehabilitation: a theoretical perspective. Behav.Brain Res. 56, 1-10.

Sulakhe, N., Elias, L.J., Lejbak, L., 2003. Hemispheric asymmetries for gap detection depend on noise type. Brain Cogn 53, 372-375.

Swisher, L., Hirsh, I.J., 1972. Brain damage and the ordering of two temporally successive stimuli. Neuropsychologia 10, 137-152.

Tallal, P., 1980. Auditory temporal perception, phonics, and reading disabilities in children. Brain Lang 9, 182-198.

Tallal, P., Gaab, N., 2006. Dynamic auditory processing, musical experience and language development. Trends Neurosci. 29, 382-390.

Tallal, P., Miller, S., Fitch, R.H., 1993. Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann.N.Y.Acad.Sci. 682, 27-47.

Tallal, P., Miller, S.L., Bedi, G., Byma, G., Wang, X., Nagarajan, S.S., Schreiner, C., Jenkins, W.M., Merzenich, M.M., 1996. Language comprehension in language-learning impaired children improved with acoustically modified speech. Science 271, 81-84.

Tallal, P., Newcombe, F., 1978. Impairment of auditory perception and language comprehension in dysphasia. Brain Lang 5, 13-34.

Tallal, P., Piercy, M., 1973a. Defects of non-verbal auditory perception in children with developmental aphasia. Nature 241, 468-469.

Tallal, P., Piercy, M., 1973b. Developmental aphasia: impaired rate of non-verbal processing as a function of sensory modality. Neuropsychologia 11, 389-398.

Tallal, P., Stark, R.E., 1981. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. J.Acoust.Soc.Am. 69, 568-574.

Treub, S., Rabinovitch, M., 1972. Auditory-linguistic sensitivity in early infancy. Developmental Psychology 6, 74-77.

Turner, R., Howseman, A., Rees, G. & Josephs, O. (1997). Functional imaging with magnetic resonance. In: Franckowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J.& Mazziota, J.C. (Eds.), Human brain functions: p 467-486. San Diego: Academic press.

Von Steinbuchel, N., 1995. Temporal systems states in speech processing. In: Herrmann, H.J., Wolf, D.E., Pöppel, E. (Eds.), Supercomputing in brain research: From tomography to neural networks. World Scientific, Singapore, pp. 75-81.

Von Steinbuchel, N., 1998. Temporal ranges of central nervous processing: clinical evidence. Exp.Brain Res. 123, 220-233.

Von Steinbuchel, N., Wittmann, M., Strasburger, H., Szelag, E., 1999. Auditory temporal-order judgement is impaired in patients with cortical lesions in posterior regions of the left hemisphere. Neurosci.Lett. 264, 168-171.

Vroom, P.A., Timmers, H., Tempelaars, S., 1977. On the hemispheric representation of time. In: Dornic, S. (Ed.), Attention and Performance VI. Erlbaum, Hillsdale, New Jersey, pp. 231-245.

Warren, J.D., Scott, S.K., Price, C.J., Griffiths, T.D., 2006. Human brain mechanisms for the early analysis of voices. Neuroimage 31, 1389-1397.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb.Cortex 11, 946-953.

Zatorre, R.J., Belin, P., Penhune, V.B., 2002a. Structure and function of auditory cortex: music and speech. Trends Cogn Sci. 6, 37-46.

Zatorre, R.J., Bouffard, M., Ahad, P., Belin, P., 2002b. Where is 'where' in the human auditory cortex? Nat.Neurosci. 5, 905-909.

2. Methodological examinations

A main drawback of conventional fMRI designs used to investigate auditory cortex functions is the acoustic noise produced during operation of the MR system, which affect the experimental integrity of auditory fMRI studies (Amaro *et al.,* 2002; Haller *et al.,* 2005; Moelker and Pattynama, 2003). At the perceptional level, recognition of auditory stimulation is distorted by acoustic masking of the stimulus when its spectral components overlap with the scanner acoustic noise as well as by psycho-acoustic effects, such as a stapedius muscle reflex that alters the perceived sound level and spectral characteristics of the presented stimulus. At the physiological level, the ambient scanner noise is tied to a saturation of the neuronal population residing in auditory cortices in the supratemporal plane. At the cognitive level the extraction of an auditory stimulus in an obscuring acoustic background is affected by foreground-background decomposition processes and attentional effort, which may provoke additional functional responses in extra-auditory frontal areas. To overcome these drawbacks of noisy fMRI it has become convenient to use scanning protocols that are reliant on the sparse sampling technique (STA) to avoid interference with ambient scanner noise. However, compared to continuous scanning these acquisition schemes require longer total imaging time in order to collect sufficient data for powerful statistical analyses. Hence, we have “silent" event-related sparse temporal acquisition designs further elaborated upon to produce clustered-sparse temporal acquisitions (CTA). In this acquisition design, multiple volume scans are acquired in rapid succession after each trial (therefore "clustered"), in order to combine advantages of "silent" scanning with time-efficient data collection.

The following two studies were conducted to evaluate and establish “clustered-sparse temporal acquisitions” protocols for auditory fMRI research. The first study investigated the effects of ambient scanner noise

bound to continuous fMRI on hemodynamic responses to spoken sentences. For this purpose, we compared data collected from a conventional “noisy”- and a clustered fMRI protocol, systematically varying the emission of acoustic scanner noise while the total amount of scanning time and auditory stimulation is kept constant. In the second study we systematically compared a clustered (CTA) and sparse temporal acquisition protocol (STA), while keeping stimulus material, task and the total duration of the experiment constant. Here two event-related auditory experiments were performed during which the subjects were presented with short tonal and auditory sentence stimuli in the context of both a CTA and a STA scheme.

2.1 Study A:

Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task

CONNY F. SCHMIDT[1,2,3], **TINO ZAEHLE**[2,5], MARTIN MEYER[2,4],
EVELINE GEISER[2], PETER BOESIGER[1], LUTZ JANCKE[2]

[1]Institute for Biomedical Engineering, University and ETH Zurich, Switzerland
[2]Department of Neuropsychology, University of Zurich, Switzerland
[3]Singapore Bioimaging Consortium (SBIC), BioMedical Sciences Institutes, Singapore
[4]Institute of Neuroradiology, University Hospital of Zurich, Switzerland

[5] Conny Schmidt and Tino Zaehle equally contributed to this manuscript

Published in:

Human Brain Mapping 2007, Feb 22; [Epub ahead of print]

Abstract

Sparse temporal acquisition schemes have been adopted to investigate the neural correlates of human audition using blood-oxygen-level dependent (BOLD) based functional magnetic resonance imaging (fMRI) devoid of ambient confounding acoustic scanner noise. These schemes have previously been extended to clustered-sparse temporal acquisition designs which record several subsequent BOLD contrast images in rapid succession in order to enhance temporal sampling efficiency.

In the present study we demonstrate that an event-related task design can effectively be combined with a clustered temporal acquisition technique in an auditory language comprehension task. The same fifteen volunteers performed two separate auditory runs which either applied customary fMRI acquisition (CA) composed of continuous scanner noise or "silent" fMRI built on a clustered temporal acquisition (CTA) protocol. In accord with our hypothesis, the CTA scheme relative to the CA protocol is accompanied by significantly stronger functional responses along the entire superior temporal plane. By contrast, the bilateral insulae engage more strongly during continuous scanning. A post-hoc region-of-interest analysis reveals cortical activation in subportions of the supratemporal plane which varies as a function of acquisition protocol. The middle part of the supratemporal plane shows a rightward asymmetry only for the CTA scheme while the posterior supratemporal plane exposes a significantly stronger leftward asymmetry during the CTA. Our findings implicate that silent fMRI is advantageous should fMRI studies unambiguously explore auditory and speech functions residing in the supratemporal plane.

Introduction

Since the advent of neuroimaging techniques approximately twenty years ago our understanding of the cerebral organization of speech functions has substantially changed and magnified. However, customary functional magnetic resonance imaging (fMRI) is hampered by acoustic noise produced during operation of the MR system, which affect the experimental integrity of auditory fMRI studies (Amaro *et al.,* 2002; Haller *et al.,* 2005; McJury and Shellock, 2000; Moelker and Pattynama, 2003). In particular, the MR acquisition inevitably produces bursts of acoustic noise up to 120 dB SPL which obscure the quality of auditory stimulus presentation. At the perceptional level recognition of spoken utterances is distorted by acoustic masking of the stimulus when its spectral components overlap with the scanner acoustic noise as well as by psycho-acoustic effects, such as a stapedius muscle reflex that alters the perceived sound level and spectral characteristics of the presented stimulus[1]. At the physiological level, the ambient scanner noise is tied to a saturation of the neuronal population residing in auditory cortices in the supratemporal plane. At the cognitive level the extraction of an auditory stimulus in an obscuring acoustic background is affected by foreground-background decomposition processes and attentional effort which may provoke additional functional responses in extra-auditory frontal areas.

This slew of deleterious issues imperatively required the development of new devices and strategies to overcome the drawback of noisy fMRI, especially in the realm of auditory cognition. Hardware based attempts to alleviate acoustic scanner noise at its very source, e.g. engineering of Lorentz force-balanced gradient coils, selection of gradient coil material with high stiffness in order to minimize mechanical vibration, implementation of acoustically damped mounting systems as well as special sound insulation have been successfully applied (Edelstein *et al.,* 2002; Katsunuma *et al.,*

[1] The stapedius reflex is a reflexive contraction of the stapedius muscle in response to loud sound. The reflex threshold varies across individuals between 70-100 dB sound pressure level in normal hearing subjects, and hence is relevant for fMRI experiments (Olsen, 1999].

2002; Mechefske *et al.,* 2002). Moreover the acoustic features of the scanner noise could be changed so that they are less disruptive (Hennel *et al.,* 1999; Seifritz *et al.,* 2006). However, neither of these approaches can be considered an optimal solution since they do not reduce acoustic noise sufficiently nor do they eliminate it. In addition, recent evaluation of one experimental design to reduce scanner noise, the *sparse temporal acquisition technique* (STA) demonstrates the advantages of this approach in auditory experiments (Amaro *et al.,* 2002; Edmister *et al.,* 1999; Gaab *et al.,* 2006a; Gaab *et al.,* 2006b; Hall *et al.,* 1999; Hall *et al.,* 2001; Talavage *et al.,* 1999; Talavage and Edmister, 2004). Using this technique, the interval between single volume scans is enhanced and the physiological delay between the onset of the stimulation and the succeeding hemodynamic response is used to separate the functional responses evoked by the scanner acoustic noise and the auditory stimulus at issue. Thus, the STA scheme allows insertion of relatively long silent intervals between function image acquisitions during which auditory stimuli are presented. Hence the aforementioned detriments affecting the perceptual, physiological, and cognitive level are reduced.

To investigate the cerebral substrates of auditory speech perception it has become convenient to use scanning protocols which are reliant on the sparse sampling technique to avoid interference with ambient scanner noise. Evidently, the functional response to an auditory stimulus as measured with a STA was magnified in terms of amplitude of the response and spatial extent of significant clusters of activation. Moreover, it has been shown that functional activation was localized with enhanced spatial and functional specificity (Yang *et al.,* 2000). However, compared to continuous scanning these acquisition schemes require longer total imaging time in order to collect sufficient data for powerful statistical analyses. Hence, “silent" event-related sparse temporal acquisition designs have been further elaborated upon to produce clustered-sparse temporal acquisitions (CTA)[2]. In this

[2] Importantly, we would like to emphasize that in the context of our study we relate the term “clustered temporal acquisition” to clustered acquisitions of scans, as opposed to the clustered volume acquisition (CVA), that is the clustered acquisition of slices within one volume scan (Edmister et al.,

acquisition design, multiple volume scans are acquired in rapid succession after each trial (therefore "clustered"), in order to combine advantages of "silent" scanning with time-efficient data collection. Incorporating long acquisition intervals, such schemes have been employed in a variety of studies, such as to investigate functional activation evoked by scanner noise itself (Bandettini *et al.,* 1998), motor activation (Bandettini *et al.,* 1998; Eden *et al.,* 1999), and in the context of short sine wave tone perception (Di Salle *et al.,* 2001) as well as in fMRI-studies on syllable processing (Zaehle *et al.,* 2004), auditory imagery (Bunzeck *et al.,* 2005), and short sentence stimuli (Rodd *et al.,* 2005; Schwarzbauer *et al.,* 2006).

1999]. The clustered sparse temporal acquisition and the continuous acquisition techniques are throughout the remainder of the paper referred to as CTA and CA techniques, respectively.

The present study

The current study investigates the effects of ambient scanner noise bound to continuous fMRI on hemodynamic responses to spoken sentences. For this purpose, we compare data collected from two differential scanning protocols that systematically vary the emission of acoustic scanner noise while the total amount of scanning time and auditory stimulation is kept constant. As outlined above scanner noise affects auditory processing at the perceptual, physiological, and cognitive level. In terms of the perceptual level, noise is meant to acoustically mask stimulus recognition. Hence we hypothesise that behavioural performance is slower and less accurate during "noisy" CA relative to "silent" CTA. As for the physiological level, persistent acoustic scanner noise comes along with a steady saturation of neural ensembles in the auditory cortex and hence reduces the functional modulation range with which the cortex can respond to an auditory stimulus. Thus, we hypothesize that in the presence of continuous scanner noise the functional response to an auditory stimulus is reduced in primary and secondary auditory areas whereas cortical association areas supporting higher linguistic processing are not affected. Furthermore, the presence of acoustic scanner noise demands additional cognitive processing. Thence, we predict a stronger involvement of extra-auditory, in particular frontal brain regions when participants undergo the experimental task in the noisy environment.

Materials and Methods

Subjects

Fifteen healthy volunteers (8 males, 7 females, age 26 ± SD 5 years), all consistently right-handed as determined by the Annett-Handedness-Questionnaire (Annett, 1992; Jancke, 1996), participated in the study. All participants were native speakers of German and had not been familiar with the stimuli prior to scanning. They had no neurological or psychiatric history, nor did they have any hearing disorder. Written informed consent

was obtained prior to the examination. The study was in accordance with Zurich Medical Faculty Ethical guidelines.

Stimuli and Design

Stimulus material consisted of normal and pseudo sentences of 3.4 ± 0.4 s duration. In the normal speech condition participants heard normal German sentences while the pseudo speech condition was composed of grammatically correct nonsense sentences with all content words having been replaced by pseudowords adhering to the phonotactical rules of the German language (for details see (Friederici *et al.,* 2000)). We decided to include normal and pseudosentences to offer participants a larger variability of controlled speech material.

An exemplary sentence from the normal speech condition was:

Die besorgte Mutter sucht das weinende Kind.

The anxious mother searches for the crying child.

Likewise an example sentence from the pseudo speech condition was:

Das mumpfige Folofel hongert das apoldige Trekon.

The mumpfy folofel hongers the apoldish trekon.

In both normal and pseudo speech condition half the stimuli were active voice sentences and half were passive voice sentences. We controlled all normal and pseudo sentences for syntax, mean duration, and mean amplitude. All sound files were digitized at a 16 bit/44.1 kHz sampling rate and were volume balanced using Volume Balancer software (Version 1.3, www.delback.co.uk/volbal/).

Task

Throughout the experiment participants were asked to indicate whether each sentence had an active or passive syntactic structure and gave their response by button press (only using the right hand) as soon as they identified the sentence structure. Prior to scanning participants underwent a

brief training during which they were presented with spoken normal and pseudo sentences specifically recorded for practice purposes.

Procedure

During scanning the room lights were dimmed and a fixation cross was projected via a forward projection system onto a translucent screen placed at the end of the magnet's gurney. Subjects viewed the screen through a double mirror attached to the head coil**.** Stimuli were controlled using Presentation® software (Version 0.70, www.neurobs.com). Stimulus presentation was synchronized with the data acquisition by a 5 V TTL trigger pulse. We used an MR-compatible piezoelectric auditory stimulation system incorporated into standard Philips headphones for binaural stimulus delivery.

Each volunteer performed two experimental runs (CA and CTA). We balanced the order of acquisition scheme across the subjects. A total of 60 auditory events and 20 empty trials (silence) was presented in each experimental run in an event-related pseudo-randomized order. Each trial lasted 15 s, leading to a total duration of 20 minutes for both the CA and the CTA scheme. Therefore, the total amount of scanning time was kept constant between the two acquisition paradigms. Each sentence was only presented once either during CA or CTA. In the CA we varied the inter-stimulus-interval systematically by moving the onset of the stimulation in three steps of 0.5 s (cf. Figure 1) (Dale, 1999). In the CTA the stimulus was presented 4.0 or 4.5 s prior to the data acquisition. Each trial was initialized visually by a 2 s fixation cross, which directly preceded the auditory stimulus or empty trial. The sound level of the acoustic noise produced by the MR scanner during the CA scheme was approximately 97.1 dB. Considering the level of attenuation provided by the combination of headphones (Phillips Standard SHC headset, attenuation of approximately 15 – 17 dB) and earplugs (attenuation of approximately 30 – 40 dB) participants were exposed to a noise level of approximately 40 - 52 dB during the CA scheme.

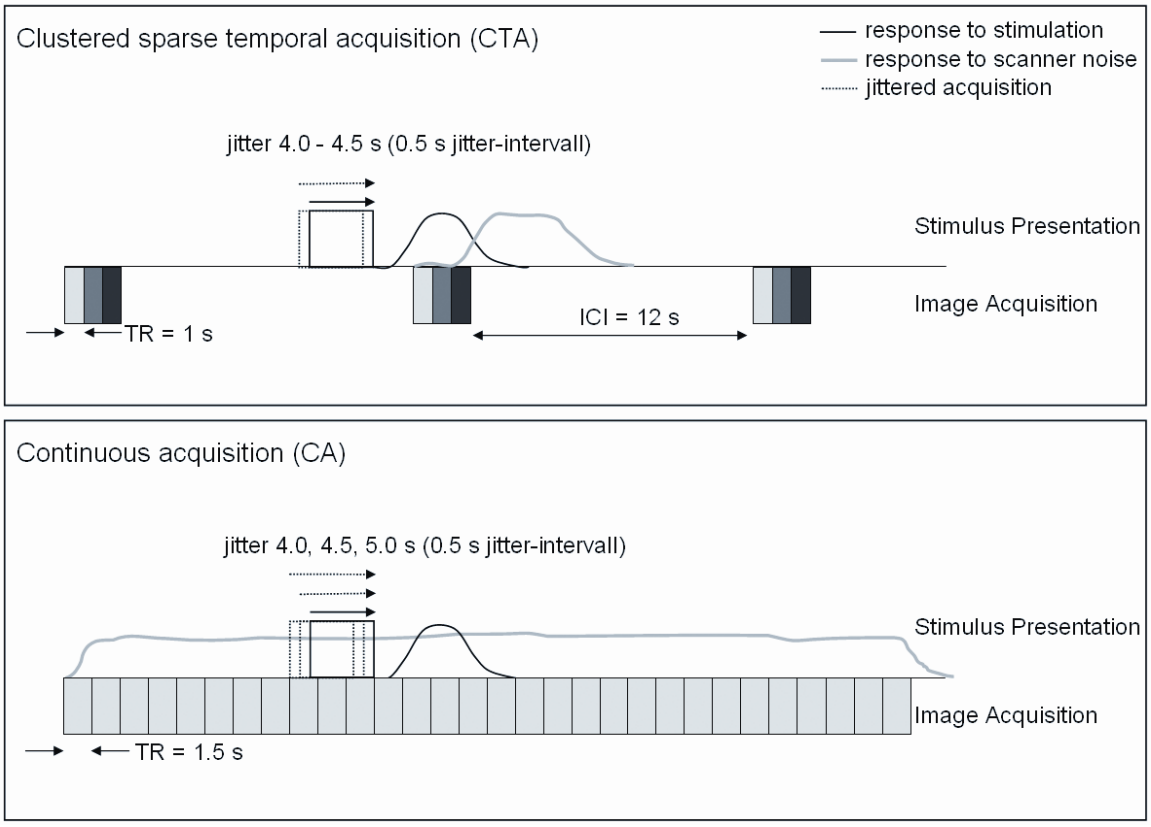

Figure 1: Clustered temporal and continuous acquisition (CTA and CA). With the CTA the auditory stimulus was presented during silence, followed by three volume scans (each with a TR = 1 s). Trial duration was 15 s with a 12 inter-cluster-interval (ICI). Stimulus presentation started 4.0 s or 4.5 s. The stimulus induced functional signal (thin black line) was sampled at the presumed amplitude of the response and was separated from the scanner-noise induced functional response (thick grey line). During continuous data acquisition volume scans were acquired with a TR = 1.5 s. The stimulus presentation was jittered by 0 s, 0.5 s and 1 s respectively, forming a 0.5 jitter interval.

Data Acquisition

Measurements were performed on a Philips Achieva 3 T whole body MR unit (Philips Medical Systems, Best, The Netherlands) equipped with an eight-channel Philips SENSE head coil. Functional data were obtained from 14 transverse slices covering the entire peri-sylvian cortex parallel to the AC-PC plane with high spatial resolution of 1.7 x 1.7 x 3 mm^3 using a Sensitivity Encoded (SENSE, Pruessmann *et al.*, 1999) single-shot gradient-echo planar sequence (acquisition matrix 128 x 128, SENSE acceleration factor R = 2.7, FOV = 220 mm, inter-slice gap 0.3 mm, TE = 35 ms). With the CTA scheme, three subsequent volumes were acquired per trial with a TR = 1 s, flip angle

90° (decay sampling), and a 12 s inter-cluster-interval (ICI) resulting in 240 dynamic scan volumes.

With the CA scheme, a total of 800 dynamic scan volumes were continuously acquired with a TR = 1.5 s (flip angle 72°). Three dummy scans preceded the experiment to allow the magnetization to reach a steady state.

We additionally collected a standard 3D T1-weighted scan for anatomical reference with 1 x 1 x 0.8 mm^3 spatial resolution (acquisition matrix 224 x 224, TE = 2.30 ms, TR = 20 ms, flip angle 20°).

Data Analysis

Artifact elimination and image analysis was performed using MATLAB 6.5 (Mathworks Inc., Natick, MA, USA), the SPM99 software package (Institute of Neurology, London, UK, http://www.fil.ion.ucl.ac.uk) and the MarsBar toolbox for SPM (http://marsbar.sourceforge.net/) (Brett *et al.,* 2002). All volumes were realigned to the first volume to account for movement artifacts, normalized into standard stereotactic space (voxel size 2 x 2 x 2 mm^3, template provided by the Montreal Neurological Institute), and smoothed using a Gaussian kernel with 8 mm full-width-at-half-maximum. Condition and subject effects were estimated using the General Linear Model (GLM) (Friston *et al.,* 1995a; Friston *et al.,* 1995b). For CA-scheme, condition of interest corresponding to the verbal condition was modeled using the appropriate stimulus function (box-car) convolved with a canonical hemodynamic response function (HRF). Low-frequency drifts were removed using a temporal high-pass filter (cut-off of 210 seconds). Additionally, the time series was temporally smoothed with a moving Gaussian kernel (FWHM = 4 seconds). Similar, for the CTA – scheme, condition of interest corresponding to the verbal conditions were modeled using a box-car function without convolving with the canonical haemodynamic response function as reference waveform. Low-frequency drifts were removed using a temporal high-pass filter (cut-off of 42 s). Because of the very low sampling frequency in this sparse temporal sampling design, low-pass filter was not applied. One characteristic of the CTA design is that the longitudinal magnetization does not reach its steady state within the acquisition of one

cluster. Thus, each of the three scans collected for one trial is influenced by different T1 saturation effects. To account for T1-decay related changes in the MR-signal, we included two additional regressors of no interest into the single subject analyses. These regressors model the means of different MR signal intensities at the three time points of acquisition along the T1-decay curve. Considering a constant amount of noise within the signal of these images and no T1-effects on BOLD-induced signal changes, we can assume a constant Contrast-to-Noise-Ration (CNR) over the three consecutive images. Thus, the use of regressors should be considered an appropriate device to partial out the influence of T1-decay related signal changes. After estimation of model parameters for each subject, an analysis of variance was calculated for the whole group using the individual contrast images for the main effects. To test hypotheses about regionally specific condition effects, linear contrasts were employed in the context of a random effects procedure (Friston *et al.,* 1999). In a first step, we calculated the effects of auditory stimulation separately for the CA and CTA- scheme. In a next step, we directly compared the data of the CA- and CTA schemes. Results of all analyses were thresholded at T = 5.52 (P = 0.05, corrected for multiple comparisons) and a spatial extent of k = 20 voxels.

Region of Interest Analysis

In order to quantify functional hemispheric differences of local brain activity in auditory fields stretching along the entire supratemporal plane (STP) as a function of the acquisition technique, a set of six distinct regions of interest (ROI) was placed for each subject in Heschl's gyrus (HG, middle STP), the planum polare (PP, anterior STP) and planum temporale (PT, posterior STP) of each hemisphere. Regions of interest were anatomically defined based on macroanatomical landmarks (Rademacher *et al.,* 2001; Steinmetz *et al.,* 1990). The left and right HG ROI were centred at MNI coordinates x = +/-49, y = -15, z = 5, size 4.16 cm^3, the left and right PP ROI at x = +/-46, y = -4, z = -7, size 3.22 cm^3 and the PT ROI comprised 3.22 cm^3 at x = +/-54, y = -30, z = 12. Mean beta-values were collected for each subject from each ROI (HG, PP, PT), and acquisition technique (CTA and CA),

and were subjected to a repeated-measure within-subjects (2 x 3 x 2) ANOVA with factors acquisition scheme (CTA, CA), ROI (HG, PP, PT), and hemisphere (left, right). All main effects or interactions with two or more degrees of freedom in the numerator were adjusted due to the Greenhouse-Geiser correction. Subsequently, we computed a (2x2) ANOVA with the factors design and hemisphere for each ROI. The threshold for significance was set at $p < 0.05$.

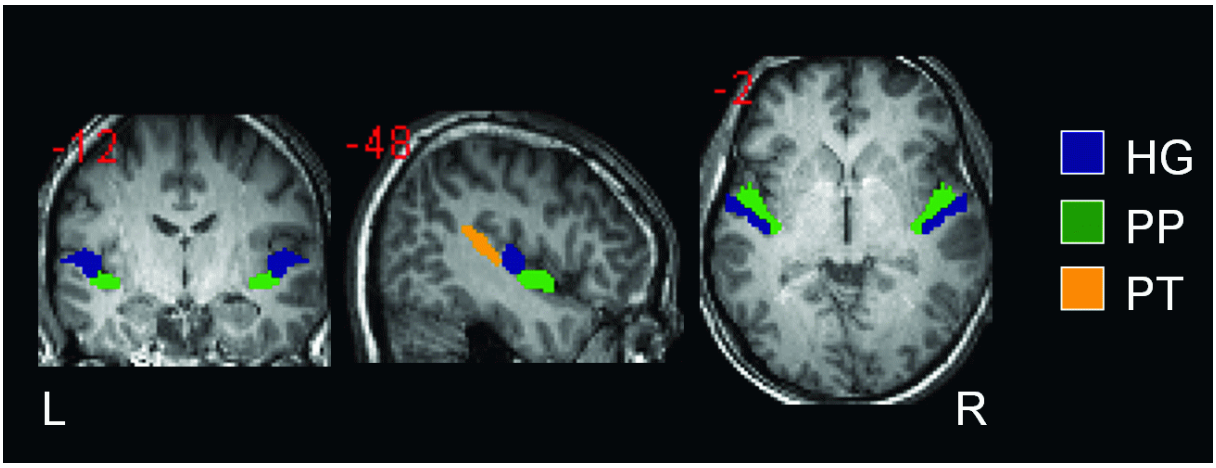

Figure 2. Figure shows the locations of ROIs. Regions of interest were anatomically defined based on macroanatomical landmarks as described by Steinmetz, H., Rademacher, J., Jancke, L., Huang, Y.X., Thron, A. & Zilles, K. (1990). Total surface of temporoparietal intrasylvian cortex: diverging left-right asymmetries. Brain Lang, 39, 357-372.

Results

Behavioural Data

The mean accuracies with which the subjects identified active and passive sentence structures in the presented auditory sentence material were 98 % and did not differ between CTA and CA. As for the reaction times, a paired sampled t-test uncovered that mean reaction times were significantly faster during "silent" clustered (RT = 2.98 ± SE 0.73 s) than during continuous scanning (RT = 3.22 ± SE 0.73 s; $t(14) = -3.74$, $p < 0.005$).

Imaging data

Akin to former studies using the same stimulus corpus (Friederici *et al.,* 2000; Heinke *et al.,* 2004; Meyer *et al.,* 2002) we collected significant functional activation in response to auditory speech stimuli[3] as compared to silence from bilateral peri-sylvian regions, in particular along the supratemporal plane, including Heschl's gyrus (HG), the plana temporale (PT) and polare (PP), from the lateral and deep fronto-opercular region (FOR), and from the insula regardless of the selected design. Furthermore, we also observed activation in extra-sylvian brain sites, namely the middle temporal gyrus, the cuneus, and subcortical structures (basal ganglia, thalamus).

To exemplify the individual acquisition main effects Figure 3 shows data of all 15 volunteers.·In terms of the CTA scheme, hearing sentences compared to empty trials evoked considerably stronger responses in bilateral superior temporal and inferior frontal regions relative to the CA scheme.

As shown in Figure 4 and listed in Table 1, the direct comparison of the CTA and CA scheme showed that the former corresponds to more significant activation bilaterally in superior temporal cortex and in the left lateral convexity of inferior frontal gyrus. For the complementary comparison of CA scheme with the CTA scheme we did not find significant activation with the given threshold. By applying a lower threshold (T = 3.30, P = 0.001, uncorrected for multiple comparisons, extent threshold k = 20 Voxels), the analysis exposed an involvement of the left and right insula and the superior occipital cortex for the CA in comparison to CTA scheme.

[3] We did not look at the contrast "normal" with "pseudosentences" (or vice versa) as it is beyond the scope of this study. This contrast and its related neural and psycholinguistic implications has been extensively addressed by former publications of the corresponding author ((Friederici et al., 2000; Heinke et al., 2004; Meyer et al., 2002; Meyer et al., 2003]) or others (Mazoyer et al., 1993; Roder et al., 2002]. We simply included pseudosentences for reasons of a highly desirable variability of speech input.

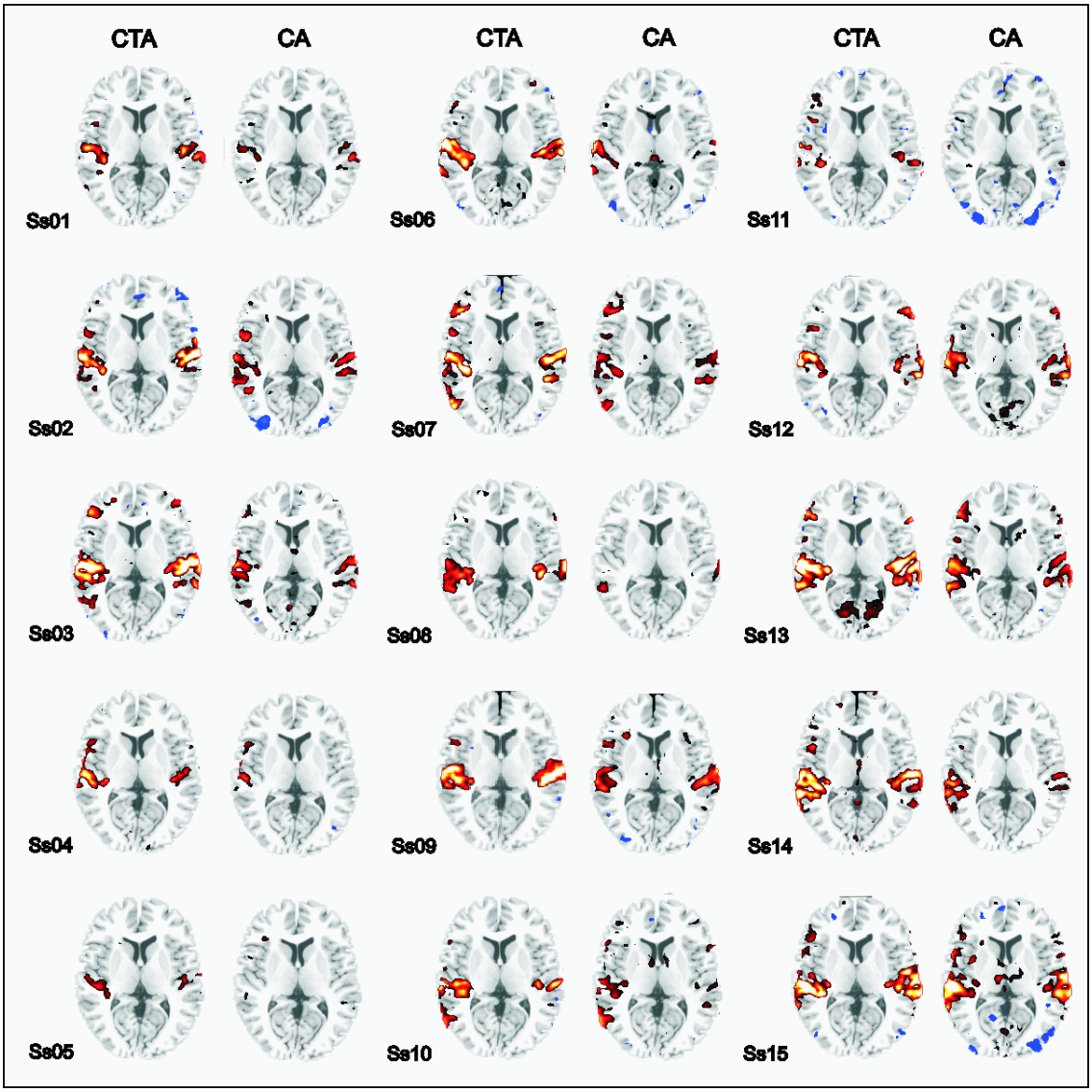

Figure 3: Individual acquisition main effects. Illustration of main effect (CA vs. CTA) in six individuals. Stronger responses to CTA are colour-coded at a red-yellow scale. Starker responses to CA are colour-coded at a blue-petrol scale. Individual functional results are superimposed onto a horizontal section (z = 8) of a standard anatomical template and are thresholded at T = 5.03 (P = 0.05, corrected for multiple comparisons). Data is plotted in neurological convention.

ROI analysis

We computed an additional ROI analysis to reveal local differences and interaction effects between acquisition scheme, ROI, and hemisphere in the supratemporal response pattern.

Table 1: Activations observed for the comparison of CTA vs. CA.

Condition and anatomical labeling	**L/R**	***T*-value**	**x**	**y**	**z**
CTA vs. CA **					
Superior temporal gyrus	R	14.35	44	-18	2
	R	13.98	38	-24	10
Superior temporal gyrus	L	13.33	-36	-28	10
	L	13.12	-38	-36	8
Lingual Gyrus	R	8.91	12	-72	4
	R	8.54	18	-66	4
Inferior Frontal Gyrus	L	8.84	-52	24	-4
	L	7.97	-48	34	10
CA vs. CTA *					
Insula	L	5.41	-36	2	12
	L	5.01	-42	6	0
Superior occipital gyrus	L	5.21	-38	-86	18
Superior occipital gyrus	R	5.21	48	-78	14
Insula	R	4.16	40	6	0

T-scores and coordinates of clusters according to the MNI stereotactic space are indicated. To assess the significance of an activation focus, results were thresholded with (**) T = 5.52, (P = 0.05 corrected for multiple comparisons) and an extent threshold of k = 20 voxels; and (*) T = 3.30 (P = 0.001 uncorrected) for multiple comparisons) and an extent threshold: k = 20 voxels).

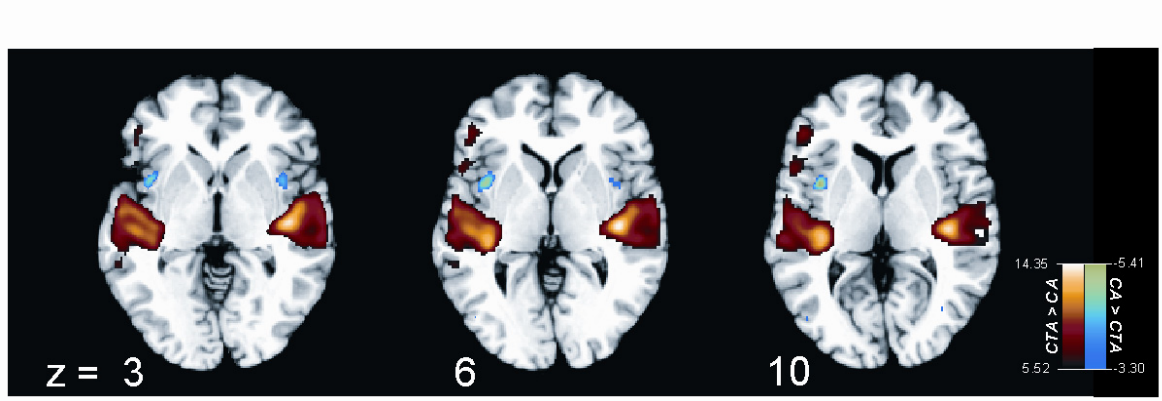

Figure 3: Group acquisition main effects (CA vs. CTA). Stronger responses to CTA are colour-coded at a red-yellow scale. Starker responses to CA are colour-coded at a blue-petrol scale. Results are superimposed onto horizontal sections of a standard anatomical template. Data is plotted in neurological convention.

A global 2 x 3 x 2 ANOVA with factors acquisition scheme (CTA, CA), ROI (HG, PP, PT), and hemisphere (left, right) revealed a significant main effect of acquisition scheme ($F(1,14) = 153.16$, $p < 0.000$) reflecting stronger responses for the CTA scheme in all regions; hemisphere ($F(1,14) = 8.59$, $p < 0.01$) indicating a general leftward asymmetry, and ROI ($F(2,28) = 13.59$, $p < 0.001$). Furthermore, we identified an interaction between ROI and hemisphere ($F(2,28) = 15.02$, $p < 0.000$), that reflects differential functional asymmetries within the ROIs, a significant interaction between acquisition scheme and ROI ($F(2,28) = 17.64$, $p < 0.000$), and an interaction of acquisition scheme by ROI by hemisphere ($F(2,28) = 14.39$, $p < 0.000$).

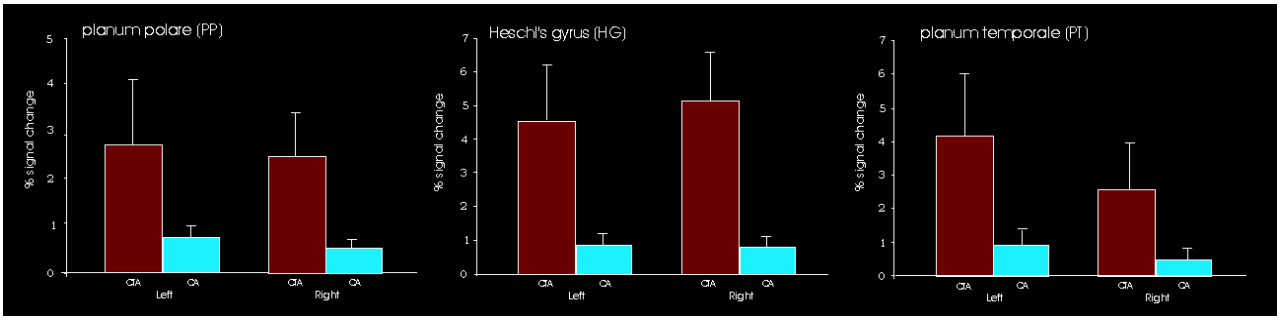

Figure 5. Analysis of three bilateral superior temporal plane regions for the main effects of acquisition scheme (CA vs. CTA) and hemisphere. Mean beta-values averaged across all subjects within bilateral planum polare (PP), Heschl's gyrus (HG), and planum temporale (PT). Error bars indicate the standard deviation. L = left, R = right.

To test the statistical significance of regional difference in functional lateralization we performed a 2 x 2 ANOVA with factors acquisition technique (CA, CTA) and hemisphere (left, right) for each ROI separately. The pattern of results for PP, HG, and PT turns out differently: The 2 x 2 ANOVA for the PP revealed a main effect of acquisition technique ($F(1,14) = 56.85$, $p < 0.000$), but no main effect of hemisphere, and no interaction between acquisition type and hemisphere. The 2 x 2 ANOVA for the HG uncovered a main effect of acquisition technique ($F(1,14) = 152.66$, $p < 0.000$), no main effect of hemisphere, and a trend for the interaction of acquisition type with hemisphere ($F(1,14) = 4.51$, $p = 0.052$) indicating a rightward functional superiority for the CTA exclusively. The 2 x 2 ANOVA for the PT unveiled a main effect of acquisition technique ($F(1,14) = 59.61$, $p < 0.000$), a main effect of hemisphere ($F(1,14) = 35.07$, $p < 0.000$), and a interaction between acquisition type and hemisphere ($F(1,14) = 35.07$, $p < 0.000$) that demonstrates a stronger leftward asymmetry for CTA relative to CA.

Generally, as apparent from Figure 4 we noted consistently larger responses in all ROIs for the CTA as for the CA scheme regardless of hemisphere. More specifically, for the PT the leftward asymmetry was larger for the CTA relative to CA scheme while for the mid STP (HG) we observed a rightward asymmetry for the CTA only.

Taken together, the fMRI results indicate that auditory sentence processing during "silent" CTA generally corresponded to more salient functional responses in auditory fields and inferior frontal sites. The CA, however, appears to lead to stronger brain responses in bilateral insula regions.

Discussion

In this study we compared a clustered "silent" acquisition with a continuous fMRI acquisition in order to test the potential advantage of fMRI in a "silent" environment in the context of an auditory speech experiment. We observed that activation in perisylvian regions corresponding to speech perception varied as a function of either the "noisy" CA or the "silent" CTA

approach. While auditory sentence processing in a silent fMRI paradigm brought on stronger functional responses of the bilateral supratemporal plane accommodating the entire auditory cortex, the presentation of the identical auditory stimulus material in the context of a traditional continuous acquisition more strongly recruited the bilateral insulae. We also noted that the participants' behavioural performance significantly slowed down once they were tested in a noisy scanning environment.

Generally, our results are consistent with previous studies having demonstrated the adverse effects of acoustic scanner noise, most strikingly in auditory fields (Bandettini *et al.,* 1998; Bilecen *et al.,* 1998; Gaab *et al.,* 2006a; Gaab *et al.,* 2006b; Loenneker *et al.,* 2001; Shah *et al.,* 1999; Talavage *et al.,* 1999). Basically, the essence of these observations suggests that at the physiological level signal detection in the context of fMRI research is harmed especially in primary and secondary auditory cortices as continuous gradient noise itself constitutes a persistent acoustic stimulus which evokes a saturation of the neuronal population residing in the supratemporal plane. Furthermore, noisy continuous fMRI increases the BOLD baseline activation in auditory regions which yields smaller BOLD responses to sound stimuli. As we hypothesized, our present study demonstrates that this reduced functional modulation range determines a decreased hemodynamic response to spoken sentences in the entire auditory cortex. Additionally, our analysis yields that scanner noise modulates the functional hemisphere asymmetry in the mid and posterior STP differentially.

Electrophysiological studies provided evidence for the view that a noisy environment asymmetrically affects neuronal functions in the context of auditory and speech processing. An MEG study on sentence perception reported that noise presented with 75 dB sound intensity disrupted early auditory processing only in the right hemisphere (Herrmann *et al.,* 2000). Our results underline these findings in that we noticed stronger functional activation in the right relative to the left mid portion of the auditory cortex for the "silent" as opposed to the "noisy" fMRI. However, it should be outlined that right auditory cortex functioning appears to be more complex. According

to a combined fMRI and MEG study using scanner noise as auditory stimulus, the right auditory cortex shows higher sensitivity to environmental noise relative to its contralateral homolog (Mathiak *et al.,* 2002). In other words, a noisy environment more strongly interferes with proper functioning of the right auditory cortex. In addition, in terms of speech perception, it has been shown that in the presence of background scanner noise syllables and speech stimuli are typically perceived as relatively flat sounds due to spectral overlap of typical MR gradient noise and spoken language (Giraud *et al.,* 2000). This effect might alter the perceived spectral characteristics of the stimulus (Hall *et al.,* 2001) and therefore affects the functioning of the core auditory cortex.

By contrast, the present study further shows that the most posterior compartment of the STP, the planum temporale, displayed a general functional leftward asymmetry for speech perception. More importantly, this leftward asymmetry was significantly stronger when speech was processed under the silent scanning condition. As apparent from recent imaging studies the left PT appears to be preferentially driven by brief phonological cues which are considered the acoustic foundation of speech (Jancke *et al.,* 2002; Meyer *et al.,* 2005a; Zaehle *et al.,* 2004). Generally, speech processing corresponds to a functional leftward asymmetry in the human brain (Hickok and Poeppel, 2004; Meyer *et al.,* 2005b; Price *et al.,* 2005; Scott and Johnsrude, 2003; Vigneau *et al.,* 2006), in particular in the PT (Josse *et al.,* 2003). Here we show that this leftward asymmetry of the posterior supratemporal plane is significantly magnified if "silent" fMRI is applied.

Finally, the contrast of CTA vs. CA also revealed stronger activation in the left lateral inferior frontal gyrus (pars triangularis and pars opercularis) for the former acquisition (cf. Figure 3). Certainly, this finding comes as no surprise as these particular brain regions have long been considered part of the perisylvian "core language system" (MacSweeney *et al.,* 2002) and have often been reported in the context of sentence comprehension regardless of modality (Friederici *et al.,* 2006a; Friederici *et al.,* 2006b). Since some researchers proposed a preference of these regions for linguistic information (Embick *et al.,* 2000; Suzuki and Sakai, 2003) it is plausible to find the left

lateral inferior frontal gyrus involved when subjects are able to process auditory sentences properly in the absence of acoustic scanner noise. However, based on the present experiment we are not able to determine whether responses of left pars triangularis and pars opercularis to auditory sentences reflect the comprehension of grammatical information or represents less specific principles of cognitive processing that are also relevant in linguistic contexts (Bornkessel *et al.,* 2005; Grewe *et al.,* 2005; Grewe *et al.,* 2006).

Our comparison between the two acquisition techniques furthermore revealed stronger responses to continuous scanning, particularly in the bilateral insula. A number of recent imaging studies have observed the anterior insula and the adjacent deep frontal operculum as supporting auditory perception, mainly in the context of demanding top-down processing (Binder *et al.,* 2004; Halpern *et al.,* 2004; Jancke and Shah, 2002; Lewis *et al.,* 2004; Wong *et al.,* 2004; Zatorre *et al.,* 1994). Additionally it has been shown that bilateral damage to the anterior insulae may result in total auditory agnosia (Habib *et al.,* 1995) or in deficient processing of temporal auditory cues (Bamiou *et al.,* 2006). The authors of the latter study suggest that this region is an integral component of the central auditory pathway. Even though it is widely accepted that this region is (partly) essential for auditory integration and complex sound modulation, its specific functional role still remains elusive (Bamiou *et al.,* 2003). The present study reports stronger responses of the insula in the context of noisy fMRI. Hence we conjecture that the insular activation particularly observed in the CA mode indicates effortful auditory perception hampered by persistent acoustic scanner noise. Similar observation has been published by Wong et al. (2004, p. 9159) who argued that “the addition of noise introduced an extra cognitive–perception demand (i.e., signal-to-noise extraction)”, which accounts for insular activation. Furthermore, a recent fMRI study demonstrated insula activity in more distractive conventional fMRI relative to a less disruptive scanning protocol (Haller *et al.,* 2005). A similar scenario can be found in the context of our study: since enhanced effort was necessary in the noisy scanning environment to retrieve intelligible

information from the auditory stimuli, it is plausible that this particular region was significantly more engaged during continuous scanning. This observation is additionally supported by the behavioural data, which showed a significantly faster response in "silent" CTA than during continuous scanning.

The present study shows that the CTA scheme can successfully be applied to map the functional response to auditory sentence stimuli in a silent environment. At the perceptual level, faster response times during CTA can be taken as evidence for facilitated auditory sentence processing. Notably, participants performed almost faultless under the two conditions. Evidently, a ceiling effect can be considered the reason we could not find a difference in mean accuracy between the two acquisition schemes. At the physiological level we show that the CTA is generally associated with stronger functional activation in the auditory cortices as this approach is not limited by saturation of the neuronal population evoked by persistent scanner noise. Additionally, we noticed noise dependent modulations of functional hemisphere asymmetry in the mid and posterior STP.

Furthermore, insular regions more strongly responding during CA acquisition may reflect unspecific effort due to compromised perception of inflowing auditory input additionally in a noisy environment.

Conclusion

The current study explores an fMRI acquisition scheme that offers the feasibility to deliver auditory stimuli in a silent environment while it allows the acquisition of multiple images to collect sufficient samples in order to efficiently measure the hemodynamic response and thus, limits the total duration of the scanning session to a reasonable duration. Hence, the CTA scheme we have evaluated combines fMRI scanning devoid of interference with scanner acoustic noise and time-efficient data collection.

Acknowledgments

During the preparation of the manuscript, CS and MM were supported by the Swiss National Science Foundation grant 46234101. EG and TZ are supported by the Swiss National Science Foundation grant 46234103. The authors wish to thank for the continuing support of Philips Medical systems and the financial support by ETH Zurich (Strategic Excellence Project TH 7/02-02). Furthermore we are indepted to the Max Planck Institute of Human Cognitive and Brain Sciences Leipzig for providing the stimulus corpus. We also gratefully acknowledge Betina Schnepf and Simon Baumann for motivating discussions.

Reference List

Amaro, E., Williams, S.C., Shergill, S.S., Fu, C.H., MacSweeney, M., Picchioni, M.M., Brammer, M.J., McGuire, P.K., (2002): Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. J.Magn Reson.Imaging 16: 497-510.

Annett, M., (1992): Five tests of hand skill. Cortex 28: 583-600.

Bamiou, D.E., Musiek, F.E., Luxon, L.M., (2003): The insula (Island of Reil) and its role in auditory processing. Literature review. Brain Res.Brain Res.Rev. 42: 143-154.

Bamiou, D.E., Musiek, F.E., Stow, I., Stevens, J., Cipolotti, L., Brown, M.M., Luxon, L.M., (2006): Auditory temporal processing deficits in patients with insular stroke. Neurology 67: 614-619.

Bandettini, P.A., Jesmanowicz, A., Van, K.J., Birn, R.M., Hyde, J.S., (1998): Functional MRI of brain activation induced by scanner acoustic noise. Magn Reson.Med. 39: 410-416.

Bilecen, D., Radu, E.W., Scheffler, K., (1998): The MR tomograph as a sound generator: fMRI tool for the investigation of the auditory cortex. Magn Reson.Med. 40: 934-937.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., (2004): Neural correlates of sensory and decision processes in auditory object identification. Nat.Neurosci. 7: 295-301.

Bornkessel, I., Zysset, S., Friederici, A.D., von Cramon, D.Y., Schlesewsky, M., (2005): Who did what to whom? The neural basis of argument hierarchies during language comprehension. Neuroimage 26: 221-233.

Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B., 2002. Region of interest analysis using an SPM toolbox. NeuroImage 16 (2) (Abstract; Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2 –6, 2002, Sendai, Japan; http://marsbar.sourceforge.net/).

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H.J., Jancke, L., (2005): Scanning silence: Mental imagery of complex sounds. Neuroimage.

Dale, A.M., (1999): Optimal experimental design for event-related fMRI. Hum Brain Mapp. 8: 109-14.

Di Salle, F., Formisano, E., Seifritz, E., Linden, D.E., Scheffler, K., Saulino, C., Tedeschi, G., Zanella, F.E., Pepino, A., Goebel, R., Marciano, E., (2001): Functional fields in human auditory cortex revealed by time-resolved fMRI without interference of EPI noise. Neuroimage 13: 328-338.

Edelstein, W.A., Hedeen, R.A., Mallozzi, R.P., El-Hamamsy, S.A., Ackermann, R.A., Havens, T.J., (2002): Making MRI quieter. Magn Reson.Imaging 20: 155-163.

Eden, G.F., Joseph, J.E., Brown, H.E., Brown, C.P., Zeffiro, T.A., (1999): Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. Magn Reson.Med. 41: 13-20.

Edmister, W.B., Talavage, T.M., Ledden, P.J., Weisskoff, R.M., (1999): Improved auditory cortex imaging using clustered volume acquisitions. Hum.Brain Mapp. 7: 89-97.

Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., Sakai, K.L., (2000): A syntactic specialization for Broca's area. Proc.Natl.Acad.Sci.U.S.A 97: 6150-6154.

Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., Anwander, A., (2006a): The brain differentiates human and non-human grammars: functional localization and structural connectivity. Proc.Natl.Acad.Sci.U.S.A 103: 2458-2463.

Friederici, A.D., Fiebach, C.J., Schlesewsky, M., Bornkessel, I.D., von Cramon, D.Y., (2006b): Processing Linguistic Complexity and Grammaticality in the Left Frontal Cortex. Cereb.Cortex.

Friederici, A.D., Meyer, M., von Cramon, D.Y., (2000): Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. Brain Lang 74: 289-300.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., Turner, R., (1995a): Analysis of fMRI time-series revisited. Neuroimage. 2: 45-53.

Friston, K.J., Holmes, A.P., Price, C.J., Buchel, C., Worsley, K.J., (1999): Multisubject fMRI studies and conjunction analyses. Neuroimage 10: 385-396.

Friston, K.J., Holmes, A.P., Worsley, K.P., Poline, J.B., Frith, C.D., Frackowiak, R.S., (1995b): Statistical parameter maps in functional imaging: A general linear approach. Hum.Brain Mapp. 2: 189-210.

Gaab, N., Gabrieli, J. D. E., and Glover, G. H., (in press): Assessing the influence of scanner background noise on auditory processing - I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. Hum.Brain Mapp.

Gaab, N., Gabrieli, J. D. E., and Glover, G. H., (in press): Assessing the influence of scanner background noise on auditory processing - II: An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. Hum.Brain Mapp.

Giraud, A.L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., Kleinschmidt, A., (2000): Representation of the temporal envelope of sounds in the human brain. J.Neurophysiol. 84: 1588-1598.

Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D.Y., Schlesewsky, M., (2005): The emergence of the unmarked: a new perspective on the language-specific function of Broca's area. Hum.Brain Mapp. 26: 178-190.

Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D.Y., Schlesewsky, M., (2006): Linguistic prominence and Broca's area: the influence of animacy as a linearization principle. Neuroimage 32: 1395-1402.

Habib, M., Daquin, G., Milandre, L., Royere, M.L., Rey, M., Lanteri, A., Salamon, G., Khalil, R., (1995): Mutism and auditory agnosia due to bilateral insular damage--role of the insula in human communication. Neuropsychologia 33: 327-339.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., (1999): "Sparse" temporal sampling in auditory fMRI. Hum.Brain Mapp. 7: 213-223.

Hall, D.A., Haggard, M.P., Summerfield, A.Q., Akeroyd, M.A., Palmer, A.R., Bowtell, R.W., (2001): Functional magnetic resonance imaging measurements of sound-level encoding in the absence of background scanner noise. J.Acoust.Soc.Am. 109: 1559-1570.

Haller, S., Bartsch, A.J., Radue, E.W., Klarhofer, M., Seifritz, E., Scheffler, K., (2005): Effect of fMRI acoustic noise on non-auditory working memory task: comparison between continuous and pulsed sound emitting EPI. MAGMA. 18: 263-271.

Halpern, A.R., Zatorre, R.J., Bouffard, M., Johnson, J.A., (2004): Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia 42: 1281-1292.

Heinke, W., Fiebach, C.J., Schwarzbauer, C., Meyer, M., Olthoff, D., Alter, K., (2004): Sequential effects of propofol on functional brain activation induced by auditory language processing: an event-related functional magnetic resonance imaging study. Br.J.Anaesth. 92: 641-650.

Hennel, F., Girard, F., Loenneker, T., (1999): "Silent" MRI with soft gradient pulses. Magn Reson.Med. 42: 6-10.

Herrmann, C.S., Oertel, U., Wang, Y., Maess, B., Friederici, A.D., (2000): Noise affects auditory and linguistic processing differently: an MEG study. Neuroreport 11: 227-229.

Hickok, G., Poeppel, D., (2004): Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92: 67-99.

Jancke, L., (1996): The hand performance test with a modified time limit instruction enables the examination of hand performance asymmetries in adults. Percept.Mot.Skills 82: 735-738.

Jancke, L., Shah, N.J., (2002): Does dichotic listening probe temporal lobe functions? Neurology 58: 736-743.

Jancke, L., Wustenberg, T., Scheich, H., Heinze, H.J., (2002): Phonetic perception and the temporal cortex. Neuroimage. 15: 733-746.

Josse, G., Mazoyer, B., Crivello, F., Tzourio-Mazoyer, N., (2003): Left planum temporale: an anatomical marker of left hemispheric specialization for language comprehension. Brain Res.Cogn Brain Res. 18: 1-14.

Katsunuma, A., Takamori, H., Sakakura, Y., Hamamura, Y., Ogo, Y., Katayama, R., (2002): Quiet MRI with novel acoustic noise reduction. MAGMA. 13: 139-144.

Lewis, J.W., Wightman, F.L., Brefczynski, J.A., Phinney, R.E., Binder, J.R., DeYoe, E.A., (2004): Human brain regions involved in recognizing environmental sounds. Cereb.Cortex 14: 1008-1021.

Loenneker, T., Hennel, F., Ludwig, U., Hennig, J., (2001): Silent BOLD imaging. MAGMA. 13: 76-81.

MacSweeney, M., Woll, B., Campbell, R., McGuire, P.K., David, A.S., Williams, S.C., Suckling, J., Calvert, G.A., Brammer, M.J., (2002): Neural systems underlying British Sign Language and audio-visual English processing in native users. Brain 125: 1583-1593.

Mathiak, K., Rapp, A., Kircher, T.T., Grodd, W., Hertrich, I., Weiskopf, N., Lutzenberger, W., Ackermann, H., (2002): Mismatch responses to randomized gradient switching noise as reflected by fMRI and whole-head magnetoencephalography. Hum.Brain Mapp. 16: 190-195.

Mazoyer, B.M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., Mehler, J., (1993): The Cortical Representation of Speech. Journal of Cognitive Neuroscience 5: 467-479.

McJury, M., Shellock, F.G., (2000): Auditory noise associated with MR procedures: a review. J.Magn Reson.Imaging 12: 37-45.

Mechefske, C.K., Geris, R., Gati, J.S., Rutt, B.K., (2002): Acoustic noise reduction in a 4 T MRI scanner. MAGMA. 13: 172-176.

Meyer, M., Alter, K., Friederici, A., (2003): Functional MR imaging exposes differential brain responses to syntax and prosody during auditory sentence comprehension. Journal of Neurolinguistics 16: 277-300.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., (2002): FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum.Brain Mapp. 17: 73-88.

Meyer, M., Zaehle, T., Gountouna, V.E., Barron, A., Jancke, L., Turk, A., (2005a): Spectro-temporal processing during speech perception involves left posterior auditory cortex. Neuroreport 16: 1985-1989.

Meyer, M., Zysset, S., von Cramon, D.Y., Alter, K., (2005b): Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. Brain Res.Cogn Brain Res. 24: 291-306.

Moelker, A., Pattynama, P.M., (2003): Acoustic noise concerns in functional magnetic resonance imaging. Hum.Brain Mapp. 20: 123-141.

Olsen, S.O., (1999): The relationship between the uncomfortable loudness level and the acoustic reflex threshold for pure tones in normally-hearing and impaired listeners--a meta-analysis. Audiology 38: 61-68.

Price, C., Thierry, G., Griffiths, T., (2005): Speech-specific auditory processing: where is it? Trends Cogn Sci. 9: 271-276.

Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., (1999): SENSE: sensitivity encoding for fast MRI. Magn Reson.Med. 42: 952-962.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H.J., Zilles, K., (2001): Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13: 669-683.

Rodd, J.M., Davis, M.H., Johnsrude, I.S., (2005): The timecourse of neural responses to ambiguous words in sentences: An FMRI study. Journal of Cognitive Neuroscience: 76-77.

Roder, B., Stock, O., Neville, H., Bien, S., Rosler, F., (2002): Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: a functional magnetic resonance imaging study. Neuroimage 15: 1003-1014.

Schwarzbauer, C., Davis, M.H., Rodd, J.M., Johnsrude, I., (2006): Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. Neuroimage 29: 774-782.

Scott, S.K., Johnsrude, I.S., (2003): The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26: 100-107.

Seifritz, E., Di, S.F., Esposito, F., Herdener, M., Neuhoff, J.G., Scheffler, K., (2006): Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. Neuroimage 29: 1013-1022.

Shah, N.J., Jancke, L., Grosse-Ruyken, M.L., Muller-Gartner, H.W., (1999): Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. J.Magn Reson.Imaging 9: 19-25.

Steinmetz, H., Rademacher, J., Jancke, L., Huang, Y.X., Thron, A., Zilles, K., (1990): Total surface of temporoparietal intrasylvian cortex: diverging left-right asymmetries. Brain Lang 39: 357-372.

Suzuki, K., Sakai, K.L., (2003): An event-related fMRI study of explicit syntactic processing of normal/anomalous sentences in contrast to implicit syntactic processing. Cereb.Cortex 13: 517-526.

Talavage, T.M., Edmister, W.B., (2004): Nonlinearity of FMRI responses in human auditory cortex. Hum.Brain Mapp. 22: 216-228.

Talavage, T.M., Edmister, W.B., Ledden, P.J., Weisskoff, R.M., (1999): Quantitative assessment of auditory cortex responses induced by imager acoustic noise. Hum.Brain Mapp. 7: 79-88.

Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., Tzourio-Mazoyer, N., (2006): Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. Neuroimage 30: 1414-1432.

Wong, P.C., Parsons, L.M., Martinez, M., Diehl, R.L., (2004): The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. J.Neurosci. 24: 9153-9160.

Yang, Y., Engelien, A., Engelien, W., Xu, S., Stern, E., Silbersweig, D.A., (2000): A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. Magn Reson.Med. 43: 185-190.

Zaehle, T., Wustenberg, T., Meyer, M., Jancke, L., (2004): Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur.J.Neurosci. 20: 2447-2456.

Zatorre, R.J., Evans, A.C., Meyer, E., (1994): Neural mechanisms underlying melodic perception and memory for pitch. J.Neurosci. 14: 1908-1919.

2.2 Study B:

Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks

TINO ZAEHLE[1,5], CONNY F. SCHMIDT[1,2,3,6], MARTIN MEYER[4], SMON BAUMANN[5], Christof Baltes[2], PETER BOESIGER[2], LUTZ JANCKE[1]

[1]Departement of Neuropsychology, University of Zurich, Switzerland
[2]Institute for Biomedical Engineering, University of Zurich and Swiss Federal Institute of Technology Zurich, Switzerland
[3]Singapore Bioimaging Consortium (SBIC), BioMedical Sciences Institutes, Singapore
[4]Institute of Neuroradiology, University Hospital of Zurich, Switzerland
[5]School of Neurology, Neurobiology & Psychiatry,Newcastle University, United Kingdom

[5] Tino Zaehle and Conny Schmidt equally contributed to this manuscript

Published in:

NeuroImage 2007, Oct 1;37(4):1195-204.

Abstract

In the functional imaging of auditory cortical functions long silent periods between the data acquisitions prevent interferences between scanner noise and the auditory stimulus processing. Recent fMRI studies have shown that sparse temporal acquisition designs are advantageous over continuous scanning protocols on physiological, perceptual, and cognitive levels. Sparse temporal acquisition schemes (STA) which use a single volume acquisition after each trial imply the advantage of auditory stimulation devoid of ambient scanner noise but have the drawback of a reduced statistical power. To alleviate this effect STA schemes have been extended to clustered-sparse temporal acquisition (CTA) designs which record several subsequent BOLD contrast images in rapid succession.

In the present study we collected data from 13 healthy volunteers performing a speech and a tonal discrimination task using both, a CTA and STA scheme to carry out a systematic evaluation of these acquisition protocols. By statistical modeling of the fMRI data sets we revealed stronger effect sizes for the STA protocol regardless of the task, reflecting the better signal-to-noise-ratio of MR images acquired with this scheme. In contrast, we demonstrate higher statistical power for the use of a CTA protocol. Accordingly, in the context of standard fMRI analysis, the CTA protocol clearly outperformed the STA scheme at the level of single-subject analysis and fixed-effects group analysis. Our results clearly suggest that it is advantageous to acquire several sample points per trial if one wants to use the benefit of “silent” fMRI. Furthermore, our data demonstrate the feasibility of the clustered acquisition of subsequent imaging volumes along the T1-decay.

Introduction

Functional magnetic resonance imaging (fMRI) has emerged as a versatile and potent non-invasive technique for mapping human brain function. However, the acoustic noise generated by the fast switching gradients of echo-planar (EPI) sequences typically used for fMRI has particularly limited its application for research on audition and language comprehension, since it interferes with proper sensory and cognitive processing: At a perceptional level recognition of the auditory stimulus at issue is hampered by acoustic screening and psycho-acoustic effects that alter its spectral characteristics and perceived loudness. At a psychological level ambient scanner noise is tied to saturation of neuronal populations in primary and secondary auditory fields, since the scanner noise acts as a persistent auditory input in itself. Furthermore, at a cognitive level enhanced attentional effort and foreground-background decomposition processes have been demonstrated to indirectly modulate the auditory response (e.g. Shah *et al.,* 1999; Shah *et al.,* 2000; Ulmer *et al.,* 1998).

EPI acoustic noise is manifested by intense noise bursts of up to 120 dB SPL, which are generated when Lorentz forces act upon the gradient coils and eddy-currents induce mechanical vibrations in metallic structures. Hence, physical features of the MR system have been engineered and gradient pulses designed such as to reduce generation and transmission of acoustic noise (for detailed review see McJury and Shellock, 2000; Moelker and Pattynama, 2003). Yet, in order to eliminate acoustic scanner noise completely, fMRI designs which comprise long silent intervals have been advanced. These protocols exploit the physiological delay between the onset of the stimulation and the succeeding hemodynamic response. Principally, the functional data are sampled within a time window close to the vertex of the hemodynamic response to the auditory stimulus, and the acquisition window is sufficiently short to rule out that the response to the scanner noise interferes with the response to the stimulus. A subsequent, long silent period then allows both the response to the auditory stimulus and the

response to the scanner noise to cease prior to the next trial (Hall *et al.,* 1999). Several studies have shown that the functional response to an auditory stimulus measured with a silent sequence as compared to continuous scanning is enhanced in terms of the amplitude of the response and spatial specificity of the activation (Yang *et al.,* 2000). It has also been demonstrated that these sparse temporal acquisition (STA) schemes ensure that only the acoustic stimulus (but not the scanner acoustic noise) contributes to the functional response (Hall *et al.,* 1999). Furthermore, the advantage of the STA approach in auditory experiments has been clearly demonstrated in recent evaluation studies. Notably, it has been evidenced that the STA protocol displays a superiority over conventional continuous scanning protocols for auditory fMRI (Gaab *et al.,* 2006a), and that detrimental affects of the scanner-noise on the acquired BOLD-signal are caused by an interaction between increased activation level during a baseline period and nonlinearity effects during stimulation within the auditory cortex (Gaab *et al.,* 2006b).

However, the number of images per time unit is considerably smaller in sparse temporal schemes as compared to continuous scanning, and inter-scan-intervals of relatively long duration are required to disentangle the hemodynamic responses from succeeding transient single events, and hence the total scan time is extended considerably. Longer imaging time, in turn, might be paralleled by attention loss and subsequently lower functional response (Shah *et al.,* 2000). In order to combine the advantage of silent event-related sparse temporal designs with more time efficient data collection, *clustered temporal acquisition* (CTA) procedures have been employed, that essentially combine a sparse sampling scheme with a multiple volume acquisition. These clustered temporal acquisition schemes with long acquisition duration have already been applied in several studies, for example to investigate functional activation evoked by scanner noise (Bandettini *et al.,* 1998), in the context of short sine wave tone perception (Di Salle *et al.,* 2001) as well as in fMRI-studies on syllable processing (Zaehle *et al.,* 2004), and auditory imagery (Bunzeck *et al.,* 2005). In a recent evaluation study we systematically compared a CTA protocol with a

subsequent clustering of three imaging acquisitions to a conventional continuous fMRI scheme (Schmidt *et al.,* 2007). We demonstrated a clear advantage of the CTA approach in fMRI investigations on auditory language processing by uncovering significantly magnified brain responses in language related areas of the bilateral supratemporal plane for the silent acquisition scheme.

Notably, a related fMRI scheme has been described in which a set of imaging volumes were acquired in the end of a silent interval. (Rodd *et al.,* 2005; Schwarzbauer *et al.,* 2006). In contrast to the CTA protocol evaluated in the present study, the authors used trains of silent slice-selective excitation pulses during the silent periods to maintain the longitudinal magnetization in a steady-state and thus avoid a T1-related signal decay. However, here we describe a CTA protocol which utilizes the T1-decay related SNR improvement of at least the first imaging volume acquired within one cluster.

The present study

In the present study we conduct a systematic evaluative comparison between a clustered (CTA) and sparse temporal acquisition (STA), while keeping stimulus material, task and the total duration of the experiment constant. Two event-related auditory experiments are performed during which the subjects are presented with short tonal and auditory sentence stimuli in the context of both, a CTA and a STA scheme. For the CTA protocol, we subsequently acquired three imaging acquisitions after each auditory trial whereas in the STA protocol a single imaging acquisition was acquired, yielding a 3:1 ratio of sample points between the two acquisition schemes. Besides the differences regarding the number of sample points in the two acquisition protocols another important issue which differs between STA and CTA schemes pertains to the signal-to-noise ratio (SNR). Due to the long interval between the single scans without excitation pulses, in the STA protocol the full longitudinal magnetization is available for image formation resulting in a relatively high SNR. By contrast, due to the subsequent clustering of imaging acquisitions following a long interval without excitation pulses in the CTA protocol only the first out of the clustered images has a SNR that is equivalent to the single image collected in the STA scheme. For the subsequently acquired images of the cluster only a fraction of the longitudinal magnetization is available resulting in decreasing SNR, and in consequence, yielding a smaller mean SNR for the CTA relative to the STA protocol. Due to these differences in SNR between the two acquisition protocols we hypothesize stronger effect sizes for statistical modeling for the STA scheme. However, due to the larger number of sample points in the CTA scheme we predict enhanced statistical power for this protocol.

Materials and Methods

Subjects

Thirteen healthy volunteers (five males, eight females, age 27 ± SD 7), all right-handed (14 ± 1 scores according to the Chapman & Chapman questionnaire (Chapman and Chapman, 1987)), participated in the study. All

subjects were native speakers of German and were not familiar with the stimuli prior to scanning. They had no neurological or neuropsychological history, nor did they have any hearing disorder. Written informed consent was obtained prior to the examination. The study was in accordance with Zurich Medical Faculty Ethical guidelines.

Stimuli and Design

Each volunteer performed two auditory tasks which comprised either tonal (*cond1*) or auditory sentence stimuli (*cond2*) in each of two experimental runs (STA and CTA). We balanced the order of acquisition scheme across the subjects. A total of 72 auditory events (36 events for *cond1* and 36 events for *cond2*) and 44 empty trials (silence) were presented in each experimental run in a pseudo-randomized order. Each trial lasted 15 s, leading to a total duration of 29 minutes for both the STA and the CTA scheme. Therefore, the total amount of scanning time was kept constant between the two acquisition paradigms. Stimuli for cond1 consisted of sine wave, string, piano and trumpet tones, each 0.4 s long. Half the stimuli had a frequency of 349 Hz (F4, American notation) and half 466 Hz (B4, American notation). To maintain the subjects' attention throughout the experiment, they were instructed to judge whether the tones were of high or low pitch. Stimulus material for cond2 consisted of normal and pseudo sentences of 3.4 ± 0.4 s duration. Subjects heard normal German sentences and pseudo speech sentences comprising grammatically correct nonsense sentences with all content words replaced by phonotactically legal pseudo words (for details see Friederici, et al. 2000). For both normal and pseudo sentences half the stimuli were active voice sentences and half were passive voice sentences. Throughout the experiment the volunteers had to decide as to whether each sentence had an active or passive sentence structure.

All sound and speech files were digitized at a 16 bit/44.1 kHz sampling rate and were volume balanced using Volume Balancer software (Version 1.3, www.delback.co.uk/volbal/). Stimuli were controlled using Presentation® software (Version 0.70, www.neurobs.com). Stimulus presentation was synchronized by a 5 V TTL trigger pulse with the data

acquisition. We used standard Philips headphones for binaural stimulus delivery.

Data Acquisition

Measurements were performed on a Philips Intera 3 T whole body MR unit (Philips Medical Systems, Best, The Netherlands) equipped with an eight-channel Philips SENSE head coil. For both aquisition schemes functional time series were obtained from 14 transverse slices covering auditory cortex with a spatial resolution of 2.7 x 2.7 x 4 mm3 using a Sensitivity Encoded (SENSE, Pruessmann *et al.,* 1999) single-shot gradient-echo planar sequence (acquisition matrix 80 x 80, SENSE acceleration factor R = 2.0, FOV = 220 mm, TR = 1000 ms, TE = 35 ms and flip angle 90°). We used parallel imaging in order to enhance acquisition speed, allowing for 14 slice brain coverage within a TR of 1 sec. Furthermore, SENSE acceleration, or in general parallel acquisition, helps reduce geometric distortions caused by the length of the echo-planar acquisition, as well as signal loss present in deep orbitofrontal areas and around the inner ear cavities, resulting from transverse dropout when susceptibility gradients shift the MR signal outside the prescribed k-space matrix (Griswold *et al.,* 1999; Preibisch *et al.,* 2003; Schmidt *et al.,* 2005). Although parallel acquisition enhances image noise due to undersampling and non-unitary operations during image reconstruction (g factor), previous studies have shown that moderate SENSE acceleration does not significantly reduce detection power, since physiological noise in the time series poses the dominant source of noise rather than the 'image noise' (de Zwart *et al.,* 2002, for comprehensive review see also Golay *et al.,* 2004). In addition to the functional time series, we collected a standard 3D T1 weighted scan for anatomical reference with 1 x 1 x 0.8 mm3 spatial resolution (acquisition matrix 224 x 224, TE = 2.30 ms, TR = 20 ms, flip angle 20°).

Sparse temporal acquisition (STA)

For the STA scheme, one single volume was acquired during each 15 s trial, resulting in a silent inter-scan-interval of = 14 s (scan-onset asyncrony:

15 s) (cf. Figure 1). Each trial started either at 3.5, 4.5, or 5.5 s prior to the data acquisition. The total amount of functional sample points was 116 image volumes.

Clustered sparse temporal acquisition (CTA)

For the implementation of a CTA technique, we combined the principle design of a STA with a clustered acquisition of three consecutive volume scans per trial. The silent inter-scan interval was 12 s (cluster-onset asyncrony: 15 s.). Auditory stimulation started at a fixed interval of 3.5 s before acquisition of the three successive volume scans. Total number of functional sample points was 348 image volumes.

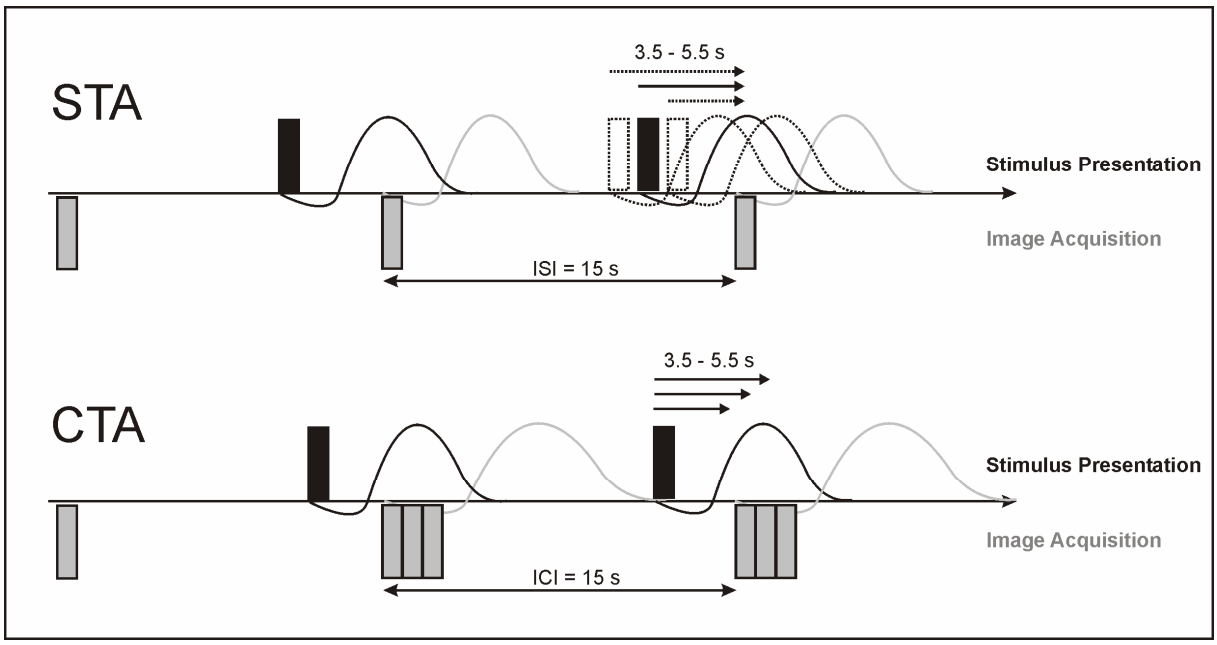

Figure 1: Sparse (STA) and clustered (CTA) temporal acquisition. With the STA the hemodynamic response (black line) is sampled only with one volume scan per trial. In the CTA the response is obtained from three consecutive volume scans.

Data Analysis

Artifact elimination and image analysis were performed using MATLAB 7.1 (Mathworks Inc., Natick, MA, USA), and the SPM99 software package (Institute of Neurology, London, UK, http://www.fil.ion.ucl.ac.uk). All volumes were realigned to the first volume to account for movement artifacts, normalized into standard stereotactic space (voxel size 3 x 3 x 3

mm3, template provided by the Montreal Neurological Institute), and smoothed using a Gaussian kernel with 8 mm full-width-at-half-maximum. Condition and subject effects were estimated using the General Linear Model (GLM) (Friston *et al.,* 1995a; Friston *et al.,* 1995b).

For the STA scheme, condition of interest corresponding to cond1 and cond2 were modeled using a box-car function without convolving with the canonical haemodynamic response function as reference waveform. Low-frequency drifts were removed using a temporal high-pass filter (cut-off of 167 s). Because of the very low sampling frequency in this sparse temporal sampling design, a low-pass filter was not applied. Similar, for the CTA scheme, condition of interest corresponding to cond1 and cond2 were modeled using a box-car function without convolving with the canonical haemodynamic response function as reference waveform. Low-frequency drifts were removed using a temporal high-pass filter and a low-pass filter was not applied.

One characteristic of the CTA design is that the longitudinal magnetization does not reach its steady state within the acquisition of one cluster. Thus, each of the three scans collected for one trial is influenced by different T1 saturation effects. To account for these T1-decay related changes in the MR-signal, we included two additional regressors of no interest into the single subject analyses. These regressors model the means of different MR signal intensities at the three time points of acquisition along the T1-decay curve. Supposing a constant amount of noise within the signal of these images and no T1-effects on BOLD-induced signal changes, we can assume a constant Contrast-to-Noise-Ration (CNR) over the three consecutive images. Thus, the use of regressors can be considered an appropriate device to partial out the influence of T1-decay related signal changes (Schmidt *et al.,* 2007).

After estimation of model parameters for each subject, we extracted the active number of voxels and the t-value of the peak voxel within the temporal cortex of each hemisphere passing a threshold of $P< 0.001$ (uncorrected for multiple comparisons), for the comparison of *cond1* and *cond2* in relation to the silent baseline. Data were submitted to a repeated-

measure within-subjects effects (2 x 2 x 2) ANOVA with the factors *acquisition scheme*, *condition*, and *hemisphere*. In a next step analyses of variance were calculated for the whole group. To test hypotheses about regionally specific condition effects, linear contrasts were employed in the context of a fixed effects (*FFX*) and a random effects (*RFX*) procedure (Friston *et al.,* 1999). For both approaches, we calculated the effects of auditory stimulation separately for the STA - and CTA – scheme. Subsequently, for each model we extracted the number of active voxels and the t-value of the peak voxels within the temporal cortex of each hemisphere passing a threshold of $P< 0.05$ (corrected for multiple comparisons). We abstained from a direct comparision of the STA and CTA data in the context of a random effects approach because the prerequisite of homoscedasticity of the two data sets is violated due to the fact that the CTA data are acquired along the T1 decay.

To further illuminate the differences between the two acquisition schemes, a region of interest analysis was performed. We used two regions of interest (ROI) covering the bilateral superior temporal cortex including Heschl's gyrus (HG, middle STP), the planum polare (PP, anterior STP), and the planum temporale (PT, posterior STP) because former studies revealed bilateral responses in these areas to the same corpus of stimuli (Baumann *et al.,* 2007; Friederici *et al.,* 2000; Schmidt *et al.,* 2007). These ROI's were anatomically defined based on macroanatomical landmarks (Rademacher *et al.,* 2001; Steinmetz *et al.,* 1990). For each individual subject averaged beta-values and averaged t-values were collected from each ROI (left, right), condition (cond1, cond2), and acquisition scheme (CTA, STA). These mean values for each ROI were subjected to a repeated-measure within-subjects effects (2 x 2 x 2) ANOVA with the factors *acquisition scheme*, *condition*, and *hemisphere*. Furthermore we collected the mean error variance (s2) for each ROI, subject and acquisition scheme by averaging the voxel values of the *ResMS*.img of each single subject model within each ROI. Subsequently, we calculated the estimated standard error (*sqrt(*s^2*c'(X'X)+*) of the effect size (c'b) within each region of interest by dividing the averaged beta-values by the averaged t-values for each ROI and subject.

Results

Single Subject Analysis

As exemplified in Figure 2, subjects displayed hemodynamic responses in the bilateral temporal cortices for both, the CTA and STA scheme. The comparison of the number of activated voxels between the STA and CTA scheme revealed a clear dominance for the CTA scheme. The ANOVA with the factors *design* (STA, CTA), *task* (cond1, cond2) and *hemisphere* (left, right) revealed a main effect of the factor *design* ($F(1,12)= 9.82$, $P<0.05$), indicating larger clusters of suprathreshold voxels for the CTA scheme, and a main effect of the factor *task* ($F(1,12)= 87.82$, $P< 0.001$), demonstrating more suprathreshold voxels for cond2. (cf. Figure 3a). Similarly, the analysis of the t-values of the peak voxel within the left and right temporal cortex for both the CTA and STA scheme revealed a main effect of the factor *design* ($F(1,12)= 27.86$, $P<0.001$), indicating significantly higher t-values for the CTA scheme, and a main effect of the factor *task* ($F(1,12)= 102.41$, $P< 0.001$), showing higher t-values for cond2 than cond1. Furthermore, the analysis showed a significant design x condition interaction ($F(1,12)= 11.09$, $P< 0.05$) (cf. Figure 3b).

In sum, the analysis showed that on the basis of the subject models, the CTA scheme clearly outperformed the STA scheme. The number of activated voxels as well as the amount of the t-value were significantly higher when hemodynamic responses during the performance of the two auditory tasks were acquired with a CTA scheme.

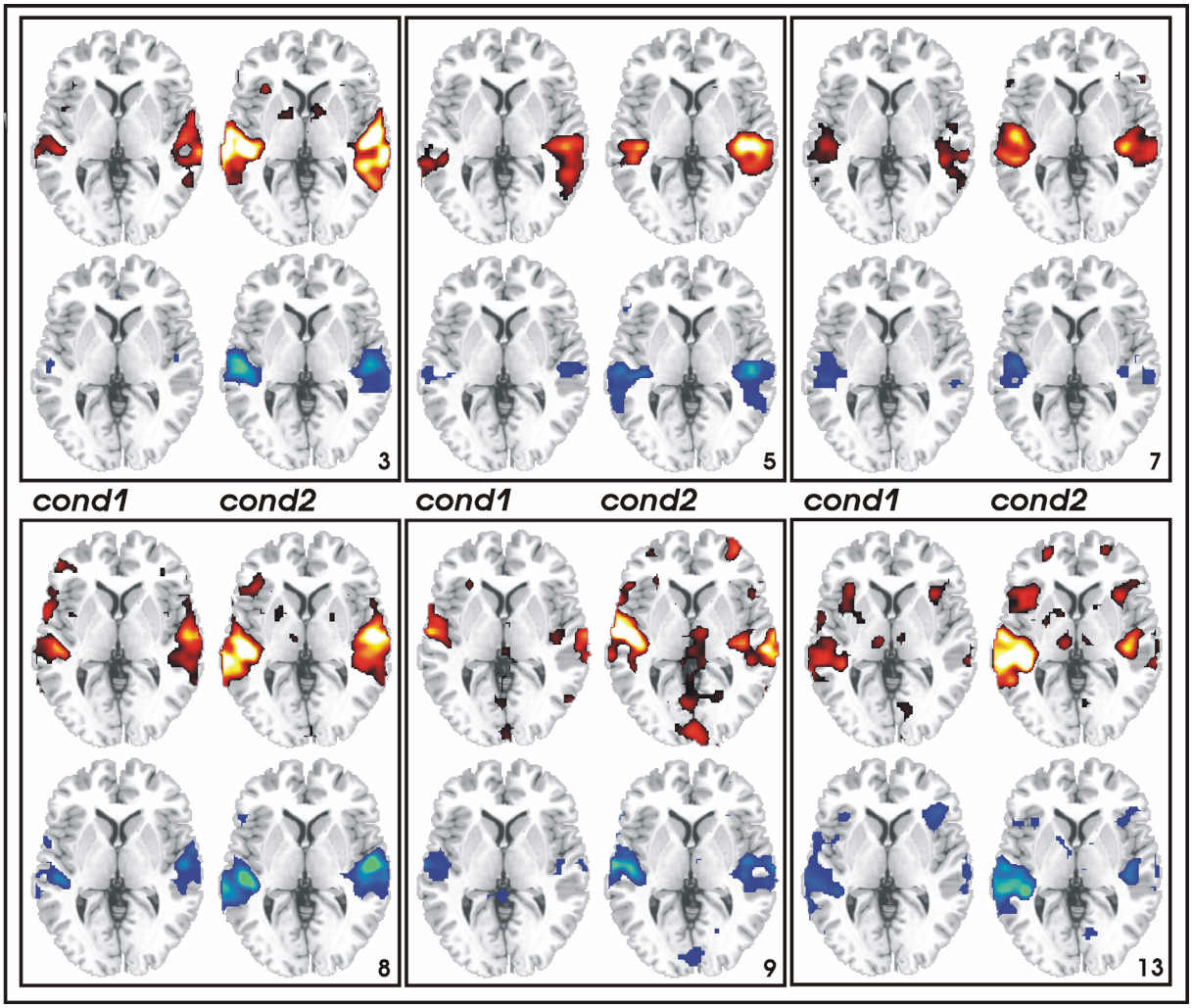

Figure 2: Individual analysis. Illustration of activations in response to auditory stimulation during *cond1* and *cond2* in six individuals. Results of the CTA scheme are colour-coded at a red-yellow scale; results of the STA scheme are colour-coded at a blue-petrol scale. Functional results are superimposed onto a horizontal section (z = 0) of a standard anatomical template and are thresholded at P = 0.001 (uncorrected for multiple comparisons). Data is plotted in neurological convention.

Group Analysis

We performed a group analysis using both a fixed effects (FFX) and a random effects (RFX) approach. For the FFX analysis, we collected significant functional activation in response to auditory stimuli as compared to silence from bilateral peri-sylvian regions along the supratemporal plane, including Heschl's gyrus (HG), the plana temporale (PT) and polare (PP), and from the lateral and deep fronto-opercular region (FOR) for both short tonal and longer speech stimuli, regardless of the acquisition scheme (cf. Figure 04). For a descriptive comparison of the FFX results of the CTA and STA

scheme, Table 01 shows the number of activated voxels and the t-value of the peak voxel within the left and right temporal cortex for both acquisition schemes. As evident from Table 01, for all contrasts the CTA scheme revealed larger clusters of suprathreshold voxels and higher t-values in comparison to the STA scheme.

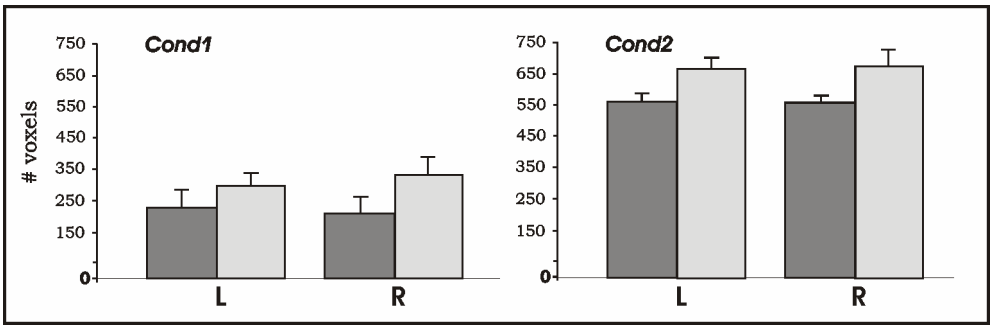

Figure 3a: Bar plots illustrate the mean number of suprathreshold voxels collected from individual analysis for the two auditory conditions (left and right panel). Each panel shows STA (dark tinge) and CTA (bright tinge) data for the left and right temporal cortex separately.

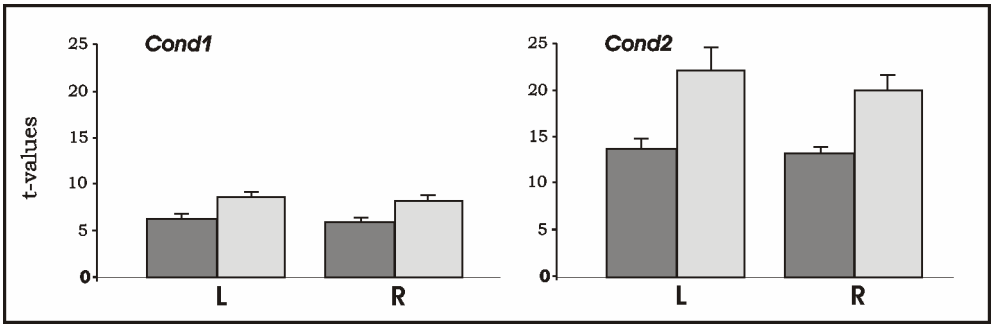

Figure 3b: Bar plots illustrate the mean T-value of the peak voxels collected from individual analysis for the two auditory conditions (left and right panel). Each panel shows STA (dark tinge) and CTA (bright tinge) data for the left and right temporal cortex separately.

Similar, for the RFX analysis, we collected significant functional activation in response to auditory stimuli as compared to silence from bilateral peri-sylvian regions along the supratemporal plane, for both short tonal (*cond1*) and longer speech stimuli (*cond2*), regardless of the acquisition scheme (cf. Figure 04). For a descriptive comparison of the RFX results of the CTA and STA scheme Table 02 shows the number of activated voxels and the t-value of the peak voxel within the left and right temporal cortex for both acquisition schemes. As apparent from Table 02, the STA scheme resulted in

slightly larger clusters of suprathreshold voxels then the CTA scheme for all performed contrasts, except the comparison of speech vs. tone stimuli. Here the CTA scheme showed larger clusters and higher t-values than the STA scheme.

In sum, the descriptive analysis of the two acquisition schemes revealed a clear dominance of the CTA scheme over the STA scheme if data are analysed with a fixed-effects approach. Contrary, if data analysis follows a random effects procedure, both designs showed comparable results with a slightly better performance of the STA scheme.

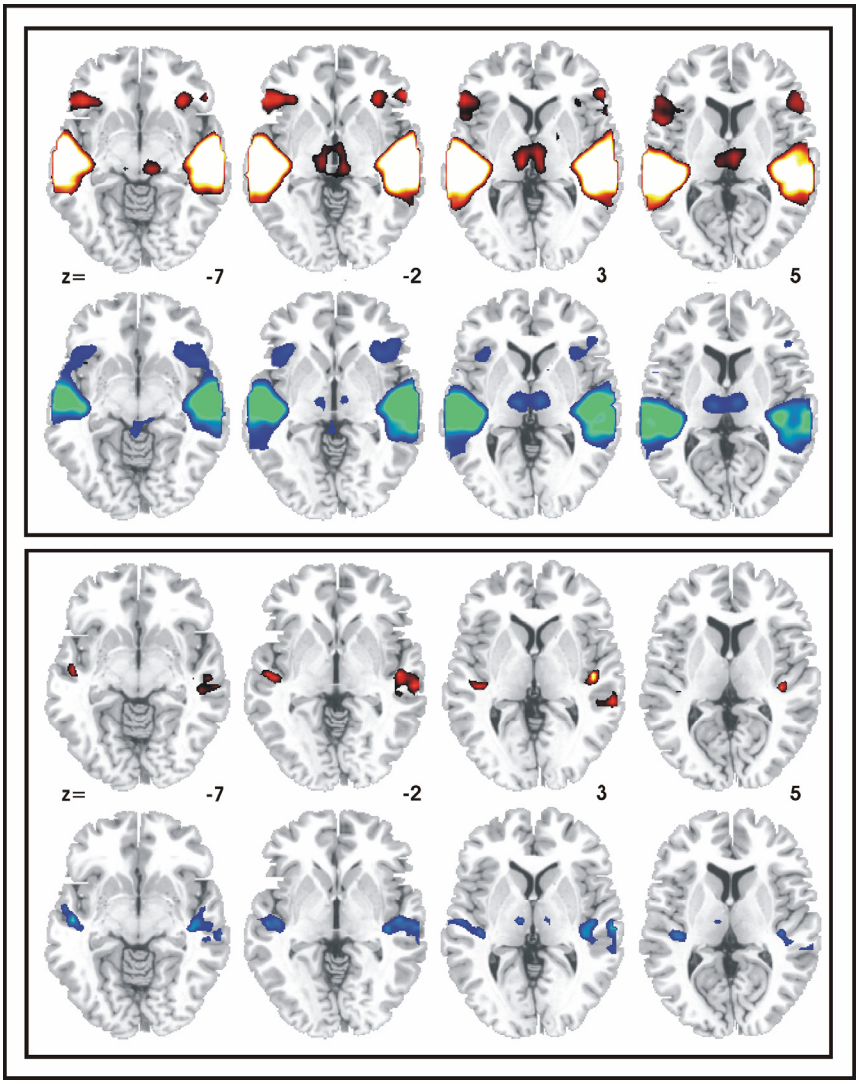

Figure 4:
Group analysis: Group activations in response to all auditory stimulation for data acquired with a CTA-scheme (colour-coded at a red-yellow scale) and a STA-scheme (colour-coded at a blue-petrol scale) separately for a fixed effects (top panel) and a random effects (bottom panel) procedure. Results are superimposed onto horizontal sections of a standard anatomical template. Data is plotted in neurological convention.

ROI Analysis

We performed the ROI analysis with two related parameters separately, namely the averaged beta-values and the averaged t-values from each individual single subject design. For the beta-values, the (2 x 2 x 2) ANOVAs with the factors *acquisition scheme* (CTA, STA), *task* (cond1, cond2), and *hemisphere* (left, right) revealed a main effect of the factor *acquisition scheme* ($F(1,12)= 5.3$ $P < 0.05$), indicating higher beta values for the STA scheme, and a main effect of *task* ($F(1,12)= 116.84$, $P<0.001$), demonstrating higher beta-values for cond2. No main effect of the factor *hemisphere* nor interactions were found (cf. Figure 5a). In contrast, for the t-values the (2 x 2 x 2) ANOVAs with factors *acquisition scheme* (CTA, STA), *task* (cond1, cond2), and *hemisphere* (left, right) revealed a main effects of the factor *acquisition scheme* ($F(1,12)= 16.63$ $P < 0.01$), indicating higher t- values for the CTA scheme, a main effect of *task* ($F(1,12)= 92.43$, $P<0.001$), demonstrating higher t-values for cond2, and a significant *scheme* by *task* interaction ($F(1,12)= 9.46$, $P<0.01$), (cf. Figure 5b).

In sum, the ROI analysis based on beta-values revealed significantly higher beta-values for the STA than for the CTA scheme and higher beta-values for the speech task (cond2) than the tone task (cond1). By contrast, the analysis based on t-values uncovered significant higher t-values for the CTA scheme than for the STA scheme. This analysis also showed significantly higher t-values for the speech task then for the tone task. Therefore, the ROI analysis using t-values confirmed the observation of the single subject analysis, in that it shows higher statistical parameters for the CTA scheme.

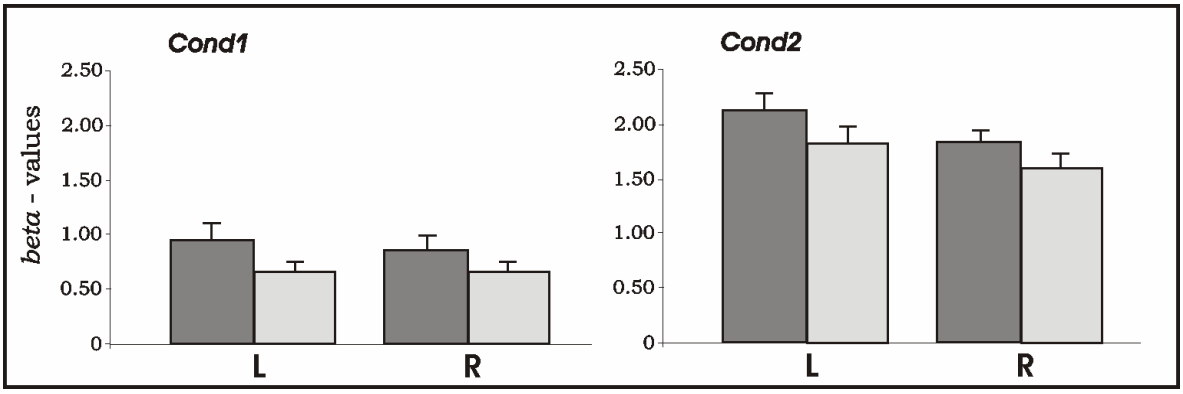

Figure 5a: ROI based on beta-value. Bar plots illustrate the mean beta-values collected from the bilateral supratemporal plane (ROI) for the two auditory conditions (left and right panel). Each panel shows STA (dark tinge) and CTA (bright tinge) data.

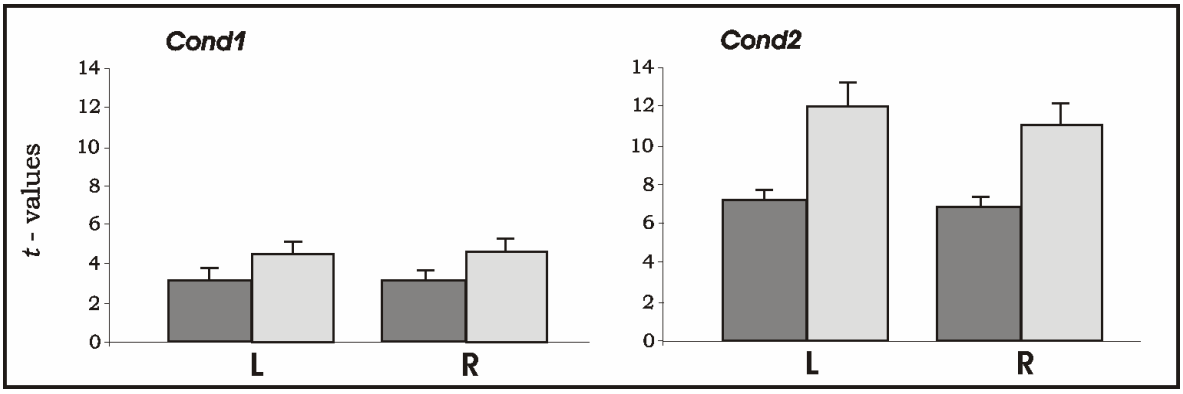

Figure 5b: ROI based on t-value. Bar plots illustrate the mean t-values collected from the bilateral supratemporal plane (ROI) for the two auditory conditions (left and right panel). Each panel shows STA (dark tinge) and CTA (bright tinge) data.

In order to explore the differences between the two acquisition protocols in more detail, we further performed ROI analyses for beta-values, t-values and error variances without separating for speech and tonal stimulation. Here we combined both conditions to an all stimulation condition. Again, we performed (2 x 2) ANOVAs with the factors *acquisition scheme* (CTA, STA), and *hemisphere* (left, right). For the beta-values, analysis revealed a main effect of *acquisition scheme* ($F(1,12)= 5.3$, $P<0.05$), confirming the effect of higher beta-values for the STA scheme. For the t-values, analysis also revealed a main effect of *acquisition scheme* ($F(1,12)= 16.89$, $P<0.001$), confirming the effect of higher t-values for the CTA scheme. Analysis of error variances (s2) revealed no significant main effect of the

factor *acquisition scheme* (F(1,12)= 1.74, P=0.21), nor a main effect of the factor *hemisphere* (F(1,12)= 1.51, P=0.24). (cf. Figure 6)

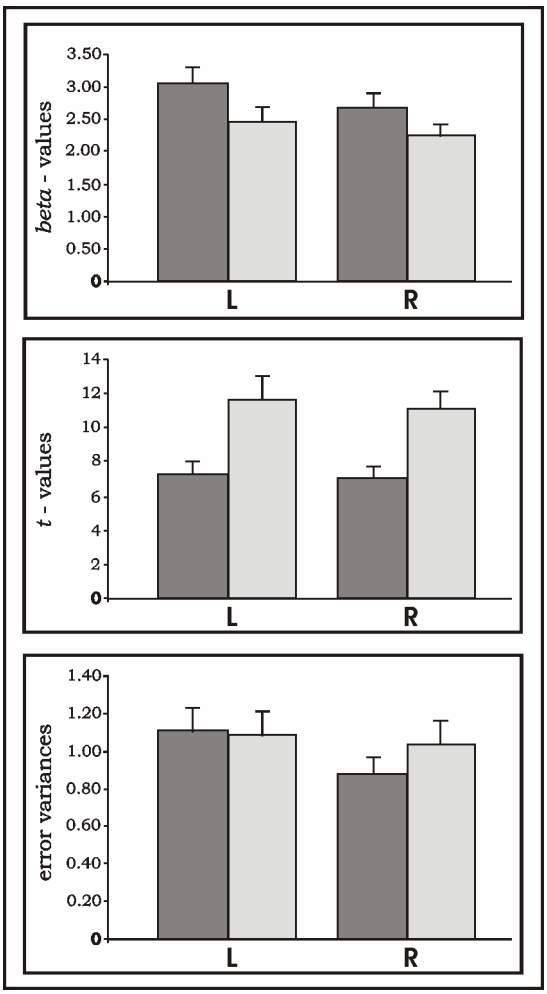

Figure 6: ROI for all stimulation condition. Bar plots denote mean beta-values (top), mean t-values (middle), and mean error variances (bottom) collected from the bilateral supratemporal plane (ROI) for the STA (dark tinge) and the CTA (bright tinge) data.

In a next step we calculated the estimated standard error (*sqrt(s²c'(X'X)+*) of the effect size (c'b) within each region of interest by dividing the averaged beta-values by the averaged t-values for each ROI and subject. Results are shown in Figure 07. The (2 x 2) ANOVAs with the factors *acquisition scheme* (CTA, STA), and *hemisphere* (left, right) revealed a significant main effect of the factor *acquisition scheme* ($F(1,12)= 116.85$, $P<0.001$), demonstrating higher averaged estimated standard errors for the STA scheme in comparison to the CTA scheme. The analysis also revealed a main effect of *hemisphere* ($F(1,12)= 5.49$, $P<0.05$), and an *acquisition scheme* x *hemisphere* interaction ($F(1,12)= 10.08$, $P<0.01$).

Furthermore, we analysed the STA/CTA ratio of the estimated standard errors. Here we revealed a value of 1.94 for the left hemisphere ROI and a value of 1.86 for the right hemisphere ROI.

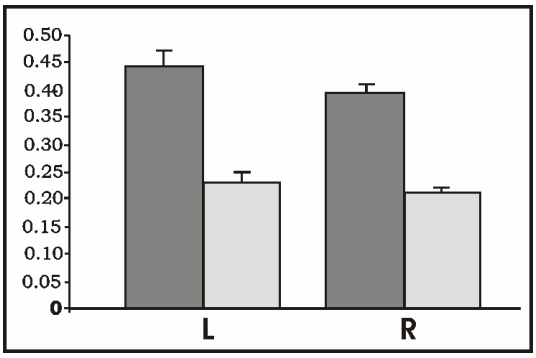

Figure 7: Plotted estimated standard deviation of parameters for all stimulation condition
Bar plots visualize mean estimated standard error for the STA (dark tinge) and the CTA (bright tinge) data for the bilateral supratemporal plane.

Discussion

In the present study we conducted a systematic comparison between a decay-sampled clustered (CTA) and a sparse temporal acquisition (STA) approach in the context of auditory cortex function. We hypothesized that the data obtained with the CTA as compared to the STA approach would exhibit enhanced statistical power, according to the larger amount of data acquired in each trial. However, T1 signal decay in the CTA attenuates the signal in the second and third volume scans and hence reduces the mean SNR within an acquisition cluster. By contrast, the STA protocol only comprises one volume scan that is not compromised by T1 signal decay

related SNR attenuation. Therefore, we expected stronger effect sizes for the STA scheme.

The systematic comparison of the two fMRI acquisition schemes revealed a clear pattern of results. At the level of single subject analyses, the CTA scheme clearly outperformed the STA scheme. In both the left and right temporal cortex the auditory stimulation related number of suprathreshold voxels was larger and the maximum t-value was significantly higher for the CTA- relative to the STA scheme. This effect was true for relative long speech and relative short tonal stimulation. However, fMRI results differentially benefited from the application of either a CTA- or STA-protocol contingent upon the statistical approach used for group analysis. Whereas in the context of a random effects approach the comparison revealed marginal differences between the two acquisition schemes, fixed effects analyses gained a tremendous benefit by the use of a CTA scheme. As revealed by the ROI analysis, these differential group results originate from contrary asymmetries in two related statistical parameters. While we observed an asymmetry in the amount of beta-values in favour of the STA scheme, an asymmetry in the height of t-values in favour of the CTA scheme became evident. By showing no differences between the error variances of the CTA and STA, we verified equal noise ratios for the two different fMRI-data sets. Furthermore, the equivalent error variances corroborate the successful application of the regressors modelling the T1-decay in the CTA scheme. The exploration of the estimated standard error of the effect sizes for CTA and STA demonstrated that the superiority of the CTA scheme originates from a reduction of the estimated standard deviation by the factor 1.94 for the left hemisphere and 1.86 for right hemisphere compared to the estimated standard error of the STA scheme. Indeed, this ratio is slightly higher as expected (SQRT(3) = 1.73).

To conclude, data show that the two protocols differ considerably with respect to their implications for auditory fMRI research. The STA scheme profits from a relatively high SNR as compared to the CTA scheme due to the full recovery of the longitudinal magnetisation prior to the acquisition of each single imaging volume. By contrast the SNR decreases within each

acquired cluster in the CTA scheme. This effect accounts for the higher beta-values brought on by the STA scheme. However, in the CTA scheme according to the higher number of sample points the predicted deviation of the estimated parameter from its true value decreases. This means, the estimation of the parameters becomes more precise and reliable, resulting in higher t-values and therefore in higher statistical power of the analysed model. Thus the CTA scheme compensates the lower mean SNR due to the three times higher number of sample points. This effect explains the higher t-values revealed by the CTA scheme and the superiority of this acquisition protocol for single-subject and fixed effects group analysis approaches.

To conclude our results clearly suggest that regardless of the stimulus duration it is advantageous to acquire several sample points per trial if one wants to use the benefit of “silent” fMRI. Furthermore, our data demonstrate that the acquisition of a cluster of subsequent imaging volumes along the T1-decay can be successfully applied in the context of auditory fMRI research.

Acknowledgments

This work was supported by Swiss National Science Foundation Grant No. 46234103 (TZ). and Swiss SNF 46234101 (CS, MM, SB). The authors also wish to thank for the continuing support of Philips Medical systems and the financial support by ETH Zurich (Strategic Excellence Project TH 7/02-02), Dr. Volker Bosch, Dr. Torsten Wüstenberg, Dr. Dorothea Weniger, and two anonymous reviewers for invaluable comments on previous drafts of the manuscript.

Reference List

Amaro, E., Williams, S.C., Shergill, S.S., Fu, C.H., MacSweeney, M., Picchioni, M.M., Brammer, M.J., McGuire, P.K., 2002. Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. J.Magn Reson.Imaging 16, 497-510.

Bandettini, P.A., Jesmanowicz, A., Van, K.J., Birn, R.M., Hyde, J.S., 1998. Functional MRI of brain activation induced by scanner acoustic noise. Magn Reson.Med. 39, 410-416.

Baumann, S., Meyer, M., Jancke, L., 2007. A short term pitch discrimination training investigated by sparsed fMRI.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H.J., Jancke, L., 2005. Scanning silence: Mental imagery of complex sounds. Neuroimage.

Chapman, L.J., Chapman, J.P., 1987. The measurement of handedness. Brain Cogn 6, 175-183.

de Zwart, J.A., van, G.P., Kellman, P., Duyn, J.H., 2002. Application of sensitivity-encoded echo-planar imaging for blood oxygen level-dependent functional brain imaging. Magn Reson.Med. 48, 1011-1020.

Di Salle, F., Formisano, E., Seifritz, E., Linden, D.E., Scheffler, K., Saulino, C., Tedeschi, G., Zanella, F.E., Pepino, A., Goebel, R., Marciano, E., 2001. Functional fields in human auditory cortex revealed by time-resolved fMRI without interference of EPI noise. Neuroimage 13, 328-338.

Edmister, W.B., Talavage, T.M., Ledden, P.J., Weisskoff, R.M., 1999. Improved auditory cortex imaging using clustered volume acquisitions. Hum.Brain Mapp. 7, 89-97.

Friederici, A.D., Meyer, M., von Cramon, D.Y., 2000. Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. Brain Lang 74, 289-300.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., Turner, R., 1995a. Analysis of fMRI time-series revisited. Neuroimage. 2, 45-53.

Friston, K.J., Holmes, A.P., Price, C.J., Buchel, C., Worsley, K.J., 1999. Multisubject fMRI studies and conjunction analyses. Neuroimage 10, 385-396.

Friston, K.J., Holmes, A.P., Worsley, K.P., Poline, J.B., Frith, C.D., Frackowiak, R.S., 1995b. Statistical parameter maps in functional imaging: A general linear approach. Hum.Brain Mapp. 2, 189-210.

Gaab, N., Gabrieli, J.D., Glover, G.H., 2006a. Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. Hum.Brain Mapp.

Gaab, N., Gabrieli, J.D., Glover, G.H., 2006b. Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. Hum.Brain Mapp.

Golay, X., de Zwart, J.A., Ho, Y.C., Sitoh, Y.Y., 2004. Parallel imaging techniques in functional MRI. Top.Magn Reson.Imaging 15, 255-265.

Griswold, M.A., Jakob, P.M., Chen, Q., Goldfarb, J.W., Manning, W.J., Edelman, R.R., Sodickson, D.K., 1999. Resolution enhancement in single-shot imaging using simultaneous acquisition of spatial harmonics (SMASH). Magn Reson.Med. 41, 1236-1245.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. Hum.Brain Mapp. 7, 213-223.

McJury, M., Shellock, F.G., 2000. Auditory noise associated with MR procedures: a review. J.Magn Reson.Imaging 12, 37-45.

Moelker, A., Pattynama, P.M., 2003. Acoustic noise concerns in functional magnetic resonance imaging. Hum.Brain Mapp. 20, 123-141.

Preibisch, C., Pilatus, U., Bunke, J., Hoogenraad, F., Zanella, F., Lanfermann, H., 2003. Functional MRI using sensitivity-encoded echo planar imaging (SENSE-EPI). Neuroimage 19, 412-421.

Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., 1999. SENSE: sensitivity encoding for fast MRI. Magn Reson.Med. 42, 952-962.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H.J., Zilles, K., 2001. Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13, 669-683.

Rodd, J.M., Davis, M.H., Johnsrude, I.S., 2005. The timecourse of neural responses to ambiguous words in sentences: An FMRI study. Journal of Cognitive Neuroscience, 76-77.

Schmidt, C.F., Degonda, N., Luechinger, R., Henke, K., Boesiger, P., 2005. Sensitivity-encoded (SENSE) echo planar fMRI at 3T in the medial temporal lobe. Neuroimage 25, 625-641.

Schmidt, C.F., Zaehle, T., Meyer, M., Geiser, E., Boesiger, P., Jancke, L., 2007. Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task. Hum.Brain Mapp.

Schwarzbauer, C., Davis, M.H., Rodd, J.M., Johnsrude, I., 2006. Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. Neuroimage 29, 774-782.

Shah, N.J., Jancke, L., Grosse-Ruyken, M.L., Muller-Gartner, H.W., 1999. Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. J.Magn Reson.Imaging 9, 19-25.

Shah, N.J., Steinhoff, S., Mirzazade, S., Zafiris, O., Grosse-Ruyken, M.L., Jancke, L., Zilles, K., 2000. The effect of sequence repeat time on auditory cortex stimulation during phonetic discrimination. Neuroimage. 12, 100-108.

Steinmetz, H., Rademacher, J., Jancke, L., Huang, Y.X., Thron, A., Zilles, K., 1990. Total surface of temporoparietal intrasylvian cortex: diverging left-right asymmetries. Brain Lang 39, 357-372.

Ulmer, J.L., Biswal, B.B., Yetkin, F.Z., Mark, L.P., Mathews, V.P., Prost, R.W., Estkowski, L.D., McAuliffe, T.L., Haughton, V.M., Daniels, D.L., 1998. Cortical activation response to acoustic echo planar scanner noise. J.Comput.Assist.Tomogr. 22, 111-119.

Yang, Y., Engelien, A., Engelien, W., Xu, S., Stern, E., Silbersweig, D.A., 2000. A silent event-related functional MRI technique for brain activation studies without interference of scanner acoustic noise. Magn Reson.Med. 43, 185-190.

Zaehle, T., Wustenberg, T., Meyer, M., Jancke, L., 2004. Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur.J.Neurosci. 20, 2447-2456.

3. Own previous related work as background and originator for further investigations

This investigation was conducted to examine the neural mechanism underlying the detection of rapid temporal auditory information using nonspeech stimuli in comparison with speech stimuli in the contex of a sparse-fMRI experiment.

Here we aimed to research to what extent the left hemisphere preferentially processes linguistically relevant temporal information available in speech and nonspeech stimuli. Furthermore, we sought to find out as to whether a left hemisphere's preference for linguistically relevant temporal information is specifically constrained to verbal utterances or if non-linguistic temporal information may also activate these areas.

3.1 Study C:

Evidence for rapid auditory perception as the foundation of speech processing – a sparse temporal sampling fMRI study

TINO ZAEHLE[1], TORSTEN WÜSTENBERG[3], MARTIN MEYER[2],

LUTZ JÄNCKE[2]

[1] Institute for Psychology, Division Cognitive Neuroscience, University of Zurich, Switzerland

[2] Institute for Psychology Division Neuropsychology, University of Zurich, Switzerland

[3] Institute of Medical Psychology, Georg-August-University Göttingen, Germany

Published in:

***European Journal of Neuroscience,* 20: 2447-2456, 2004.**

Abstract

We examined the processing of verbal and non-verbal auditory stimuli using an event-related fMRI study to reveal the neural underpinnings of rapid temporal information processing and it's relevance during speech perception. In the context of a clustered sparse-temporal fMRI data collection eight right-handed native German speakers performed (a) an auditory gap detection task and (b) a CV-syllable discrimination task. A tone perception task served as a non-temporal control condition. Here we aimed to research to what extent the left hemisphere preferentially processes linguistically relevant temporal information available in speech and nonspeech stimuli. Furthermore, we sought to find out as to whether a left hemisphere's preference for linguistically relevant temporal information is specifically constrained to verbal utterances or if non-linguistic temporal information may also activate these areas. We collected hemodynamic responses from three 'time points of acquisition' (TPA) with varying temporal distance from stimulus onset to gain an insight on the time course of auditory processing.

Results show exclusively left sided activations of primary and secondary auditory cortex associated with the perception of rapid temporal information. Furthermore, the data shows an overlap of activations evoked by nonspeech sounds and speech stimuli within primary and secondary auditory cortex of the left hemisphere. The present data clearly supports the assumption of a shared neural network for rapid temporal information processing within the auditory domain for both speech and nonspeech signals situated in left superior temporal areas.

Introduction

Recent brain imaging studies have shown that language perception is a complex function relying on the integrity of many different neural systems comprising identification of tune, intensity, human voice, voice identity, prosody, or phonetic perception recruiting both the right and the left hemisphere (Belin *et al.,* 2000, 2002; Binder *et al.,* 2000; Scott & Johnsrude, 2003; Wise *et al.,* 2001; Zatorre *et al.,* 2002). Although, these recent studies have uncovered many new findings at least one major question has been left unanswered so far. This question stems back to the late 60s and 70s of the last century and addresses the specific role the left hemisphere plays in speech perception.

Early studies on the hemispheric specialisation for speech perception using dichotic listening techniques found that volunteers exhibit a right ear advantage for most speech stimuli, indicating a greater left hemispheric processing (Bryden, 1982; Kimura, 1961; Kimura & Folb, 1968). These early findings provided strong evidence for the view that language abilities are left lateralized. Schwarz and Tallal (1980) suggested, that the right ear advantage might reflect phonetic as well as acoustic processing. They assumed that the left hemisphere is generally sensitive for rapidly changing acoustic cues irrespective of linguistic content. These rapidly changing acoustic cues, e.g. the 'Voice-Onset-Time' (VOT) are frequent characteristics available in linguistic stimuli but it is not the linguistic stimuli per se that triggers left hemispheric processing. The authors based their assumption on the finding that a deficit in the temporal information processing within the auditory system, rather than a deficit in the linguistic abilities, could account for phonological disorders observed in language-learning-impaired children (Schwartz & Tallal, 1980; Tallal & Newcombe, 1978). This hypothesis seems to be confirmed by the improvement of speech discrimination and language processing in these children after training of recognition of temporal modulated speech and non-speech sounds (Merzenich *et al.,* 1996; Tallal *et al.,* 1996). Furthermore, the combination of deficits in auditory processing of speech and non-speech sounds that

included rapid changing acoustic cues has also been shown for adult aphasics (Steinbuchel & Poppel, 1993). The co-existence of impairments at a phonological-language level and reduced discriminating abilities for temporal non-speech sounds in patient studies mentioned above indicate that the underlying deficits might be associated with inappropriate functioning of a universal cortical system mediating auditory temporal resolution in speech and nonspeech sounds.

Recent brain imaging studies have been providing evidence corroborating the clinical observations. Comparing cortical activations evoked by speech and nonspeech sounds which contains ether rapid or slow changing acoustic cues, these investigations revealed evidence for a left sided hemispheric asymmetry in primary (Belin *et al.,* 1998; Liegeois-Chauvel *et al.,* 1999) and secondary auditory areas (Jancke *et al.,* 2002; Joanisse & Gati, 2003) in response to rapidly changing acoustic cues. These findings implicate a functional association of the left hemisphere, in particular the posterior superior temporal cortex, with the processing of temporal information in speech and non-speech sounds. In the present study applying a gap-detection task we used nonspeech stimuli in comparison with speech stimuli to examine the neural mechanism underlying the detection of rapid temporal auditory information.

Methods

Subjects

Eight native speaker of German (age range: 25-38) participated in this study. After a full explanation of the nature and risks of the research, subjects gave informed consent for the study according to a protocol approved by the Ethics Committee of the Otto-von-Guericke University, Magdeburg. They had no history of any neurological, psychiatric or hearing impairment. All subjects were consistent right-handed according to a standard handedness questionnaire (Annett, 1992).

For additional testing of the hemispheric language lateralisation, we applied a fMRI language mapping task for each subject (Fernandez *et al.,* 2001). In result, all subjects showed activations exclusively in the left hemisphere. These findings were interpreted as a clear left brain lateralization of language-related processing within the examined sample.

In order to avoid gender effects in brain size and shape as well as in functional language lateralization (Luders *et al.,* 2002), only male subjects were studied. All subjects underwent 4 training sessions prior to scanning until they were familiar with the paradigm.

Stimuli

The study material comprises three experimental conditions and one control condition:

1. gap perception task: (Gap task)
2. consonant-vowel syllable perception task (CV task)
3. tone perception task (Tone task)

The three sets of auditory stimuli were generated with a sampling depth of 16 bits and a sampling rate of 44,1 kHz using the SoundForge 4.5 Software (Sonic Foundry Inc., http//www.sonicfoundry.com).

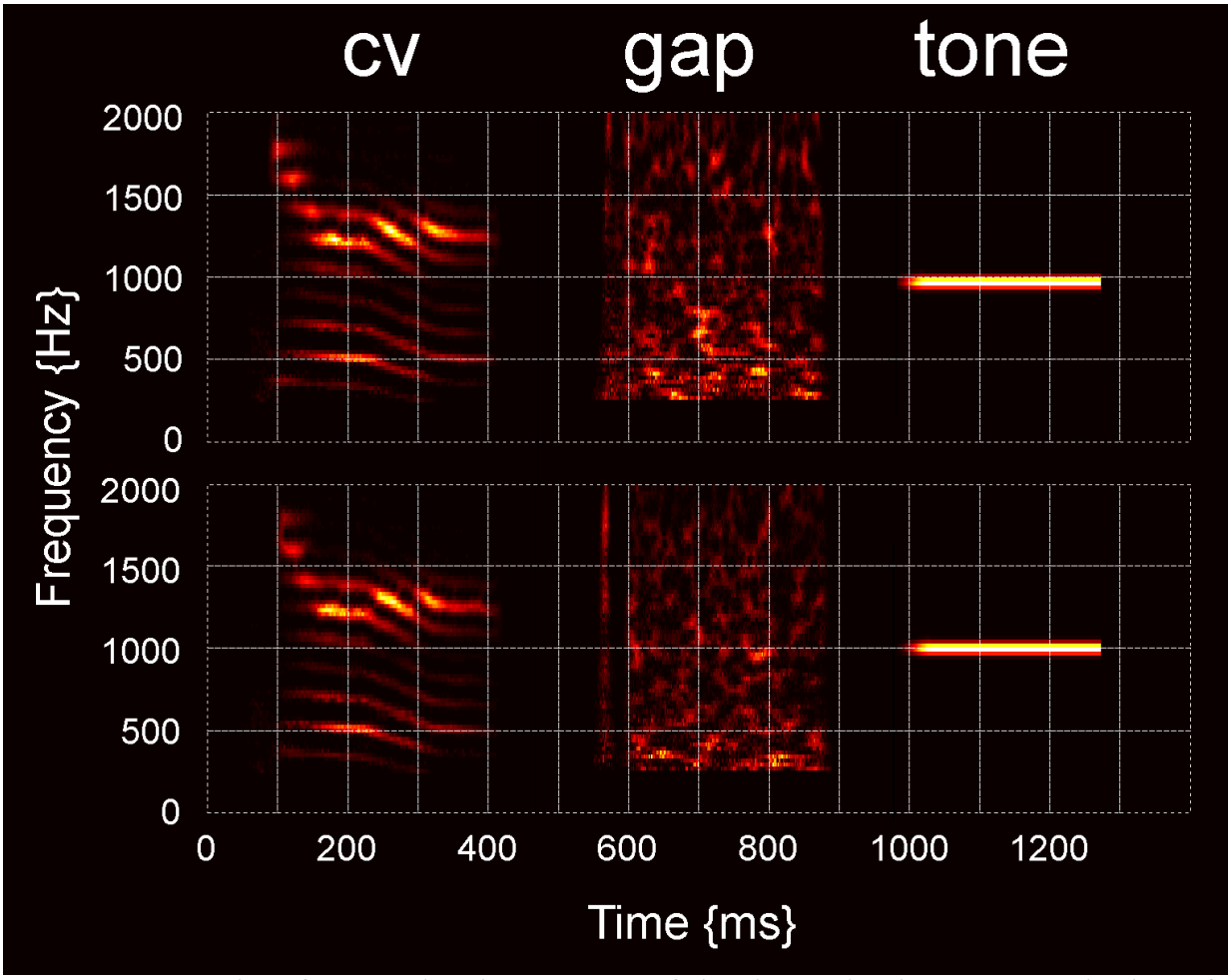

Figure 1. Examples of narrow-band sonograms of the three stimulus types used. Plots of frequency spectra as a function of time are shown. The left column illustrates CV-stimuli. The upper row shows the syllable /da/,and the lower row the syllable /ta/. The middle column shows examples of the Gap-stimuli with leading element duration of 5ms. The upper row shows the Gap-stimuli with gap duration of 8ms, and the lower row the Gap-stimuli with gap duration of 32 ms. The right column illustrates examples of the Tone stimuli with a frequency of 0.95 kHz (upper row) and 1 kHz (lower row).

The Gap detection task is the most common method used to measure auditory temporal resolution. The listener is presented with two streams of sounds, one of which has a brief silent period ("gap") at its temporal midpoint. The listener's task is to identify this signal and thus the shortest detectable gap ("gap threshold") is determined. A central gap task requires the listener to detect a gap bounded by markers (leading and trailing elements) that are different in frequency content. The gap to be detected is delimited by markers differing in spectral qualities. Performing a central gap detection task imperatively requires a relative timing of the offset of activity evoked by the leading element and the onset of activity mediating the trailing

element (Phillips *et al.,* 1997; 1998). Thus, performing gap detection needs a fine grained analysis of rapid changing acoustic cues. Therefore, for the *gap task*, we created gap stimuli according to the work of Phillips et al. (1997), containing two sound elements separated by a gap. The leading element was a wideband noise burst with a length of either 5 ms or 7 ms. The duration of the gaps was 5 ms and 32 ms. The trailing element was a bandpassed noise centred on 1.0 kHz and an width of 500 Hz. The duration of the whole stimulus was adapted for a total length of 330 ms. All noise periods had a rectangular envelope without any fading.

Using the possible four combinations of leading elements and gap duration, only such pairs of stimuli were chosen for the experiment, which had the same duration of the leading element but different gap durations (duration of leading element/gap duration : 5/5 & 5/32; 7/5 & 7/32). The order of presentation of single stimuli within the stimulus pair was balanced throughout the experiment.

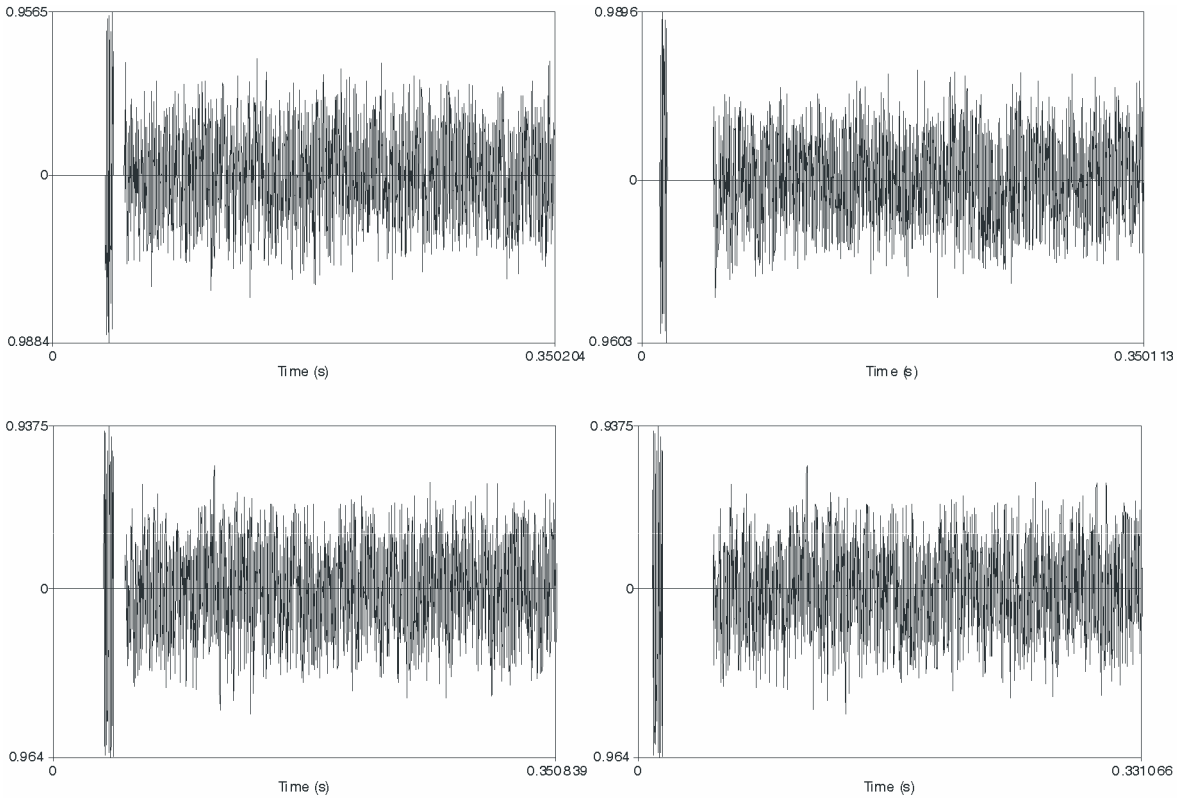

Figure. 2. Examples of wave-forms of the Gap stimuli: The upper row shows Gap stimuli with duration of the leading element of 5 ms and a duration of the gap of 8 ms (left) and 32 ms (right). The lower row illustrates Gap stimuli with a duration of the leading element of 7ms and a gap duration of 8 ms (left) and 32 ms (right).

In the *CV task* we used syllables (/ta/, /da/) recorded by a trained phonetician. The criterion for temporal alignment of the syllables was the onset of articulatory release. The duration of the syllables was 330 ms. The Voice Onset Time (VOT) of the syllables was manipulated to increase the difficulty of the CV task. (VOTs in ms for /da/ = 30, /ta/ = 40). At least stimuli were grouped into pairs consisting of both modified syllables in a randomised order.

For the stimulation in the *tone task* sine-wave tones were used. According to the mean frequencies in the gap and CV task, frequencies of 950 Hz, 1000 Hz or 1050 Hz was chosen. The tone stimuli were 330 ms long and linearly faded over the first and last 36 ms. According to the gap detection task, pairs of stimuli with difference in frequencies of 50 Hz and randomized presentation order were created.

Experimental task

Volunteers performed a delayed matching to sample task (cf Figure 3). They were asked to compare pairs of stimuli (either *gap*, *CV* or *tone*) with a probe stimulus, and press one of two response buttons to indicate which stimulus matched the probe. Each trial started with an attention cue realized by one square wave pulse (click) followed by the target stimulus and the described stimulus pair. The next trial of the same condition was repeated 2000 ms later without the attention cue. For each experimental condition, 24 trials were presented in a randomized order, separated into three runs. Forming a baseline 19 trails of a motor control task were included in each run. In this control task the subjects had to press both response buttons after perceiving one additional click instead of the gap, cv or tone stimuli. In summary the total amount of trials subjects had to performed was 162 (Subject 3 only 108 trials in 2 runs, due technical problems). There was a constant delay of 500 ms between the end of

stimulation and the onset of the first scan and a delay of 13,4 seconds between the end of the MR acquisition and the onset of the next stimulation.

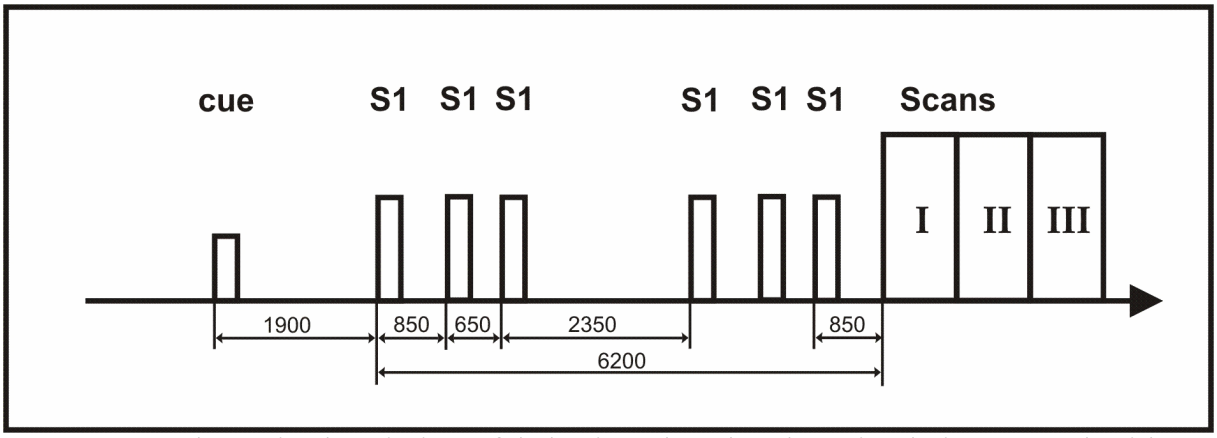

Figure. 3. Schematic description of behavioural task: The stimulation started with an acoustic cue, a square-wave pulse (click) followed 1900ms later by a stimulus (S1). This first stimulus served as the probe. The second stimulus (S2) followed 850 ms after onset of the probe and further 650ms later the third stimulus (S3) was presented. The task was to compare the two stimuli with the probe and press on of two response keys to indicate which stimuli matched the probe. After 2000 ms a new trial of the same condition started. 500ms after presentation of the last stimulus the acquisition of the three fMRI images started. The task was the same for all conditions (gap, cv, tone), only the stimulation was changed.

fMRI scanning

Functional magnetic resonance imaging was performed on a 1.5 Tesla whole-body MRI system (Signa Horizon LX, General Electrics Inc., Waukesha, WI, USA), equipped with echo planar imaging (EPI) capability using the standard circular polarized head coil for radio-frequency transmission and signal reception. Using a midsagittal scout image, 18 axial slices (slice thickness = 7 mm, inter slice gap = 1 mm) of a T2* weighed gradient echo EPI sequence were acquired parallel to the bicommissural plane and covering the whole brain. The following acquisition parameters were used: repetition time TR = 1.5 s, echo time TE = 40 ms, field of view = (200x200) mm^2, flip angle = 80°, matrix size = 64 x 64, voxel size = (3.125 x 3.125 x 8) mm^3. After collecting a cluster made up of three complete sets of slices, a time gap of 21.5 seconds without scanning followed.

During scanning the room lights were dimmed and the participants were instructed to keep their eyes close. Binaural auditory stimulation was

presented by a digital playback system including a high frequency shielded transducer system. This acoustic transmission system includes a piezoelectric loudspeaker enabling the transmission of strong sound pressure levels (V105 dB) with excellent attenuation characteristics.

fMRI data analysis

To account for different longitudinal magnetisation in the three consecutive images acquired in one cluster (and also the T1-saturation), the data were analyzed separately for each time point of acquisition (TPA). This results in three TPAs of 500-1990 ms, 2000-3490 ms and 3500-4990 ms after the offset of the last stimulus.

Artefact elimination and image analysis was performed using MATLAB 6.0 (Mathworks Inc., Natiek, MA, USA) and the SPM99 software package (Institute of Neurology, London, UK). For each TPA all images were realigned separately to the first image, spatially normalized into standard stereotaxic space ((4x4x4) mm^3, template provided by the Montreal Neurological Institute (Evans *et al.,* 1993)) and spatially smoothed using a 10 mm full-width-at-half-maximum Gaussian kernel.

Condition and subject effects were estimated using the General Linear Model (GLM) (Friston *et al.,* 1995). The effect of global differences in scan intensity was removed by scaling each scan to the global mean of all scans. Low-frequency drifts were removed using a temporal high -pass filter (cut off of 512 seconds). Because of the very low sampling frequency in this temporal sparse sampling design (1/26 Hz) low-pass filter was not applied. We used a box-car function without convolving with the canonical hemodynamic response function (HRF) as reference waveform. For each TPA, each condition and each subject these waveforms were set to “high” for auditory stimulation and “low” for the motor control task to identify regions of activations while listeners discriminated specific stimuli. After estimation of model parameters for each subject, an analysis of variance was calculated for the whole group, using the individual contrast images for the main effects. Because only the effects in the temporal lobe were of interest in this study, the statistical analysis was restricted to this part of the brain.

Therefore, a small volume correction was performed as implemented in SPM99 by using mask-images containing only the superior parts of the temporal lobe of both hemispheres (HG, PT, STG, and STS). To test hypotheses about regionally specific condition effects, linear contrasts were employed in the context of a random effects procedure. The results of these statistics were thresholded by T = 3.0 (p = 0.01 uncorrected for multiple comparisons) and k = 10 voxels. To test the hypotheses the Gap task and also the CV task were contrasted with the Tone control task (*gap>tone / cv>tone*) for each of the three different TPA.

In order to test for between-hemispheric differences the contrast images for main effects were doubled. These doubled images were horizontally "flipped", resulting in a second contrast image data set with reversed transversal orientation. By comparing the original with the flipped images hemodynamic responses obtained for the left and right hemisphere were comparable in the context of SPM analysis (Jancke *et al.,* 2002). The resulting SPM{T} for these effects were thresholded at a T value of 2.44 (P = 0.01 uncorrected for multiple comparisons) and a spatial extent criterion of k = 10 voxels. In the following we report only comparisons which pass this statistical threshold.

Following a correction for the differences in the coordinate system between the stereotaxic atlas by Talairach and Tournoux (Talairach & Tournoux, 1988) and the stereotaxic space employed by SPM99, the anatomical localization of the local maxima was assessed. Peaks located in PT and HG were additionally verified using published probability maps (Rademacher et al., 2001; Westbury et al., 1999).

Results

1. Activation pattern for the comparison between each experimental task and motor control task for three time points of acquisition (TPA).

Table 1 lists the main results obtained from comparisons between experimental conditions and motor control. Comparisons for TPA1 and TPA2 revealed strong bilateral activations of the STG for the Gap task and the CV task. For TPA3 the CV task revealed bilateral activation of the STG. The Tone task revealed bilateral activations of the STG for all TPAs.

2. Activation pattern for the comparison between the Gap task and Tone task and between the CV task and Tone task for three time points of acquisition (TPA)

Results are shown in Figure 4 and listed in Table 2. Contrasting the CV task with the Tone-task revealed bilateral activation of the STG for all three TPAs. Contrasting the Gap task with the Tone task revealed activation of the left STG for TPA1 and TPA2. No differences between the two tasks could be observed for TPA3. Referring to the probabilistic maps of the HG and PT, the subtraction of hemodynamic responses obtained during the Tone task from those obtained during the Gap task revealed a peak of activation within the left HG for TPA1. The peak of activation for the comparison of the CV task and the Tone task was found in the similar area of the left HG for TPA1. Additionally, a right sided activation increase could be observed for the CV task in relation to the Tone task within the PT and HG.

For TPA2 the comparison of hemodynamic responses for the Gap task and the Tone task showed an activation cluster with a peak in the left PT and left HG. Comparing the CV task with the Tone task revealed stronger bilateral activations of both the left as well as the right PT.

For TPA3 the Gap task in relation to the Tone task revealed no suprathreshold voxels. The comparison of the CV task and the Tone task showed peaks of activation in the right HG and left PT.

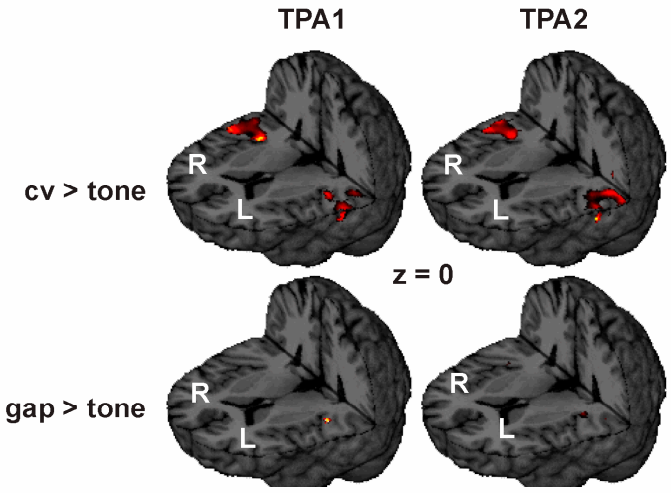

Figure. 4. Activation pattern for the comparison Gap vs. Tone and CV vs. Tone, separately for TPA1 and TPA2. SPM(t) maps are overlaid on a structural MRI brain.

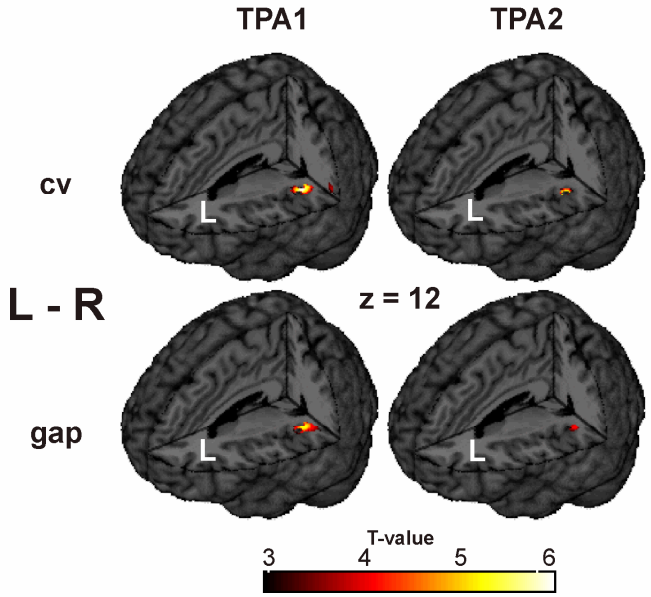

Figure. 5. Activation pattern for the direct hemispheric comparison for the Gap task and the CV task, separately for TPA1 and TPA2. SPM(t) maps are overlaid on a structural MRI brain.

Table 1: Peak Activations and its MNI Coordinates Observed for the Comparison of the Experimental Task and Motor Control Task separately for each TPA

Condition and anatomical area	Probability atlas	L/R	*k*	*T*-value	*x*,	*y*,	*z*
TPA1							
Gap > motor control							
STG	PT (26–45%)	L	418	10.57**	−54,	−18,	6
STG/HG	HG (10–20%)	R	334	10.55**	48,	−12,	3
CV > motor-control							
STG	PT (26–45%)	L	535	11.38**	−57,	−18,	6
HG	HG (70–80%)			5.16*	−39,	−30,	9
STG	PT (26–45%)	R	479	10.49**	63,	−15,	3
STG	HG (10–20%)			8.45**	48,	−21,	3
TPA2							
Gap > motor control							
STG/ HG	PT (5–25%) HG (10–20%)	L	273	8.34*	−48,	−21,	0
STG	PT (26–45%)	R	208	7.10*	63,	−9,	3
CV > motor-control							
STG	PT (5–25%)	R	426	20.28**	63,	−15,	0
HG	HG (10–20%)			9.32**	45,	−18,	6
STG	PT (26–45%)	L	532	10.08*	−66,	−18,	3
TPA3							
Gap > motor control (no suprathreshold voxels)							
CV > motor-control							
STG	PT (5–25%)	R	71	5.90*	69,	−21,	0
STG		L	59	4.36	−69,	−27,	3

The coordinates are given according to the MNI space together with its T-scores. In addition, the number of voxels passing the threshold of *P = 0.001 uncorrected for multiple comparisons and **P = 0.05 corrected for multiple comparisons are presented (k). For HG and PT, the probability that the peak voxel lay within the designated cortical region is shown according to the probability maps published for HG (Rademacher *et al.,* 2001) and PT (Westbury *et al.,* 1999).

3. Between-hemisphere differences for the Gap task and CV task

Results are shown in Figure 5 and listed in Table 2. The direct hemisphere comparison for the hemodynamic responses observed during the experimental conditions revealed an activation increase in the left PT for both the Gap task and the CV task. For TPA1 both conditions showed an overlapping cluster of activation in the left PT close to HG. For TPA2 both conditions revealed the identical location of activation peaks in the left PT. For TPA3 the right hemisphere response was stronger for the Gap task; no suprathreshold voxels could be observed for the CV task.

4. Conjunction analysis of the hemispheric comparisons for the Gap task and CV task for three TPAs

Figure 6 and Table 2 show the results of the conjunction analysis. The analysis of haemodynamic responses observed for the hemispheric differences for both the Gap task and the CV task revealed strong activational increases in the left PT for both, TPA1 and TPA2. For TPA3 the analysis revealed a cluster of activation within the right temporal lobe.

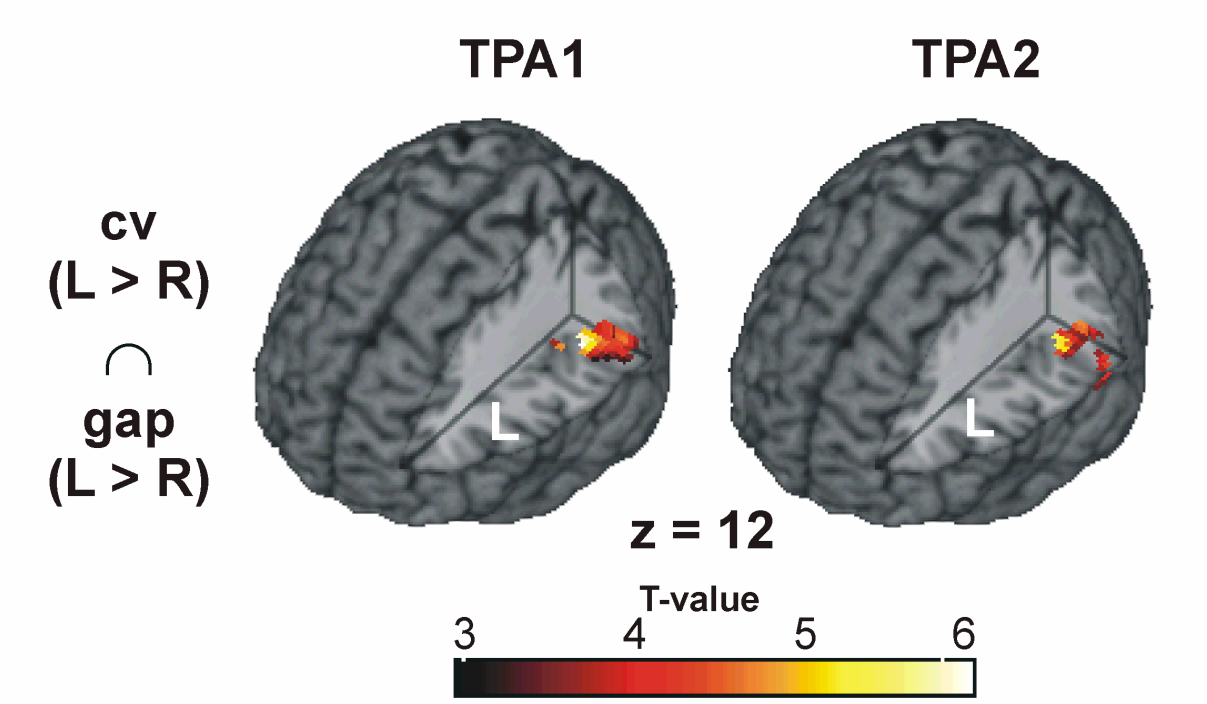

Figure. 6. Activation pattern of the conjunction analysis for the hemispheric comparisons of Gap task and CV task, separately for TPA1 and TPA2. SPM(t) maps are overlaid on a structural MRI brain.

Table 2: Peak activations observed for the Contrasts Gap Task vs. Tone Task (gap > tone), CV Task vs. Tone Task (cv > tone), and for the direct hemispheric comparison (L>R) separately for each TPA

Condition and anatomical area	Probability atlas	L/R	*k*	*T*-value	*x*,	*y*,	*z*
TPA1							
Gap > Tone							
STG/HG	HG (70–80%)	L	22	5.57*	−45,	−18,	6
CV > Tone							
STG	PT (26–45%)	R	310	8.26*	60,	−12,	−3
HG	HG (60–70%)			7.10*	42,	−18,	6
HG	HG (80–90%)	L	301	5.17*	−42,	−18,	9
STG	PT (26–45%)			5.14*	−60,	−21,	0
Gap L > R							
STG	PT (5–25%)	L	135	6.46**	−42,	−36,	18
CV L > R							
STG	PT (5–25%)	L	128	6.00**	−39,	−33,	12
HG	HG (70–80%)						
STG	PT (26–45%)	R		3.84*	57,	−45,	18
Conjunction (Gap L > R and CV L > R)							
	PT (26–45%)	L	141	5.68**	−42,	−33,	15
TPA2							
Gap > Tone							
STG	PT (46–65%)	L	20	6.85*	−51,	−36,	9
HG	HG (50–60%)			3.90	−39,	−24,	3
CV > Tone							
STG	PT (46–65%)	R	258	10.46*	66,	−27,	12
STG	PT (26–45%)	L	394	8.27*	−66,	−15,	0
Gap L > R							
STG	PT (5–25%)	L	12	2.94	−42,	−36,	21
CV L > R							
STG	PT (5–25%)	L	37	3.37*	−42,	−36,	21
STG/HG	HG (40–50%)			3.22*	−39,	−33,	12
Conjunction (Gap L > R and CV L > R)							
STG	PT (5–25%)	L	38	2.94**	−42,	−36,	21
TPA3							
Gap > Tone (no suprathreshold voxels)							
CV > Tone							
HG	HG(50–60%)	R	12	5.69*	36,	−27,	15
STG	PT (5–25%)	L	14	4.51*	−63,	−15,	0
STG		L	12	4.40	−48,	−18,	0
Gap L > R							
STG		R	19	3.45*	42,	−39,	6
CV L > R (no suprathreshold voxels)							
Conjunction (Gap L > R and CV L > R)							
		R	6	2.53*	39,	−33,	3

The coordinates are given according to the MNI space together with its T-scores. In addition, the number of voxels passing the threshold of *P = 0.001 uncorrected for multiple comparisons and **P = 0.05 corrected for multiple comparisons are presented (k). For HG and PT, the probability that the peak voxel lay within the designated cortical region is shown according to the probability maps published for HG (Rademacher *et al.*, 2001) and PT (Westbury *et al.*, 1999).

Discussion

The aim of the present study was to generally explore the neural basis for the perception of rapidly changing acoustic cues in speech and nonspeech sounds. Therefore we applied a Gap task and a CV task in the context of an fMRI setting. Even though various behavioural investigations and theoretical assumptions point to the close relationship between phonetic perception and gap detection, no neuroimaging investigation examining this matter has been done so far. Therefore by using a Gap task relative to a CV task, our study aimed at testing the hypothesis: Do acoustic nonspeech stimuli which require the processing of rapidly changing cues (short gaps) trigger activation in brain regions normally involved in phonetic perception? We designed the present fMRI study in order to identify the particular area which subserves both gap detection and phonetic perception. According to the model proposed by Tallal and Newcombe (1978) we anticipate left-biased overlapping activity in the auditory cortex during auditory gap detection and phonetic perception. As it has been demonstrated that MR-related acoustic noise interferes with fMRI-experiments in the auditory domain (Amaro *et al.,* 2002), we used a sparse temporal sampling approach optimally suited to measure hemodynamic responses in the auditory cortex (Gaab *et al.,* 2003; Hall *et al.,* 1999).

In particular, we set out in this study to address two issues 1) the functional lateralization of rapidly changing acoustic cues and 2) the fundamental role of temporal acoustic information processing during speech perception which we will be discussing in turn. We collected hemodynamic responses from three subsequent time points of acquisition (TPAs). Taken into account, that data acquisition period for TPA3 lasted from 3000 ms to 4500 ms after the onset of the scanner sound, the hemodynamic response evoked by this sound emission reaches its maximum and may interfere with stimulus induced responses. Hence, as we assumed that results of TPA3 may be confounded, only the results for TAP1 and TAP2 will be discussed.

Is there a functional lateralisation for rapid temporal information processing?

One aim of the present study was to explore cortical areas associated with the processing of gap perception. Performing a central gap detection task imperatively requires a relative timing of the offset of activity evoked by the leading element and the onset of activity mediating the trailing element. Thus, performing the Gap task needs a fine grained analysis of rapid changing acoustic cues.

Our data clearly suggests that left rather than right auditory cortex subserves the processing of rapidly changing acoustic cues. The spatial pattern of condition-related changes in hemodynamic response varies over time. Responses to Gap task when compared to Tone task were restricted to the left hemisphere. In particular, we observed a shift of activity from HG at TPA1 to PT at TPA2. Analogously, the same shift of activity in the left hemisphere could be seen for the CV task compared to the Tone task. Further evidence for the relevance of the left PT for the processing of rapidly changing cues is provided by the direct hemispheric comparison. For TPA1 and TPA2 the left PT exhibited stronger activity when directly compared to the contralateral area for both Gap and CV task. Thus, our data are generally in line with previous findings of left lateralized temporal processing within the auditory domain (Belin *et al.,* 1998; Joanisse *et al.,* 2003; Liegeois-Chauvel *et al.,* 1999).

By having analyzed brain signals collected from two consecutive TPAs we were able to delineate the temporal distribution of task related hemodynamic responses. Our data are in accordance with the assumption of a general mechanism for rapid temporal processing localized in the left HG (Liegeois-Chauvel *et al.,* 1999). In addition, our results are also in line with previous findings of selective responses observed in the PT for the perception of rapidly changing phonetic cues (Jancke *et al.,* 2002). Along the lines of this study, we also noted bilateral PT activity for the perception of phonetic cues (CV task). Unlike the phonetic task, only the left hemisphere participated in the Gap task.

Based on the converging evidence outlined above we conclude that both, the primary and secondary auditory areas of the left hemisphere are

involved in the processing of rapid temporal information. The differential activation observed for HG and PT may be explained in terms of different task demands. In the study by Liegeois-Chauvel *et al.* (1999), revealing superiority of left HG, participants had to listen passively to specific sounds whereas in the study by Jancke *et al.* (2002), exposing a superiority of left PT, participants had to categorise the syllables. Activity in the left posterior STG associated with rapid temporal auditory processing while participants performed a categorization task has also recently been reported by Joanisse *et al.* (2003). The processing of rapidly changing acoustic cues differentially engaged the HG and the PT at different time points. By integrating this evidence with the findings of the present study, we conclude that the HG is more strongly associated with analysing the sound signal per se whilst the PT may be more involved in categorisational processing which is based on the prior analysis. However, since perception of CV generally includes categorisation we cannot rule out that formerly reported regional activity differences have been biased by different imaging techniques. The present study helps clarify this issue by providing evidence that both the HG and the PT subserve the processing of rapidly changing acoustic cues.

Taken collectively, this evidence can be considered a clear support for the left hemisphere superiority in processing rapidly changing acoustic cues. Furthermore, they are in line with neuropsychological findings: Patients with left sided lesions and symptoms of fluent aphasia exhibited considerably reduced performances in auditory processing of temporal information (Robin *et al.,* 1990; Von Steinbuchel *et al.,* 1999). Additionally, research of cerebral asymmetries in neuropsychologically intact persons traditionally reports a right ear (corresponding to a left hemisphere) advantage for the discrimination of speech stimuli. Even more relevant, a right ear advantage has been reported for the perception of non-speech stimulation requiring the processing of rapid temporal information (e.g. temporal order (Mills & Rollman, 1980), duration discrimination (Mills & Rollman, 1979), discrimination offsets of tones (Emmerich *et al.,* 1981) and gap detection (Brown & Nicholls, 1997).

Is temporal information processing a fundamental component of speech perception?

Previous neuroimaging investigations have identified regions within the left and right superior temporal cortex that are involved in speech perception (Belin *et al.,* 2000; Binder *et al.,* 2000; Davis & Johnsrude, 2003; Jancke *et al.,* 2002; Meyer *et al.,* 2000; 2002; Wise *et al.,* 2001; Zatorre *et al.,* 2002). One key question raised from previous studies reads, whether these brain regions are specifically devoted to processing speech or whether they are specialized for processing several features of sounds in general. This question is closely related to the ongoing debate to what extent speech receives a full auditory analysis before it is processed as speech or whether speech is processed differently at a very early stage. It has been argued that the complexity of speech needs a unique processing mode with unique neural foundations (Liberman & Mattingly, 1985; Liberman & Whalen, 2000). Therefore a second aim of the current study was to investigate to what extent rapid temporal information processing plays a fundamental role during speech perception. The identification of areas supporting both, temporal as well as phonetic processing would lead to a better understanding of the issue of generality of speech processing. Furthermore, the notion of a common basis of speech and temporal processing would provide support for the origin and development of several speech related disorders (e.g. dyslexia). To address this question, a CV task in addition to a Gap task was performed. The Gap stimuli were designed so that they had the gross temporal properties of spoken stop consonants. They consisted of a relative short, broadband leading element as an analogue of stop consonant, followed by a brief period of silence as an analogue of the VOT, and then followed by a relative long, steady-state, low-frequency element as an analogue of a vowel. In terms of processing the voice onset times, the task is to recognize the relative timing of the high frequency consonantal burst and the low frequency vowel. Therefore, it has been assumed that performing a gap-detection task requires the same perceptual mechanisms that are also involved in stop consonant discrimination. It has been shown that for short leading elements (5ms), the gap-threshold is around 30ms (Phillips *et al.,*

1997), which is close to the VOT on the perceptual boundary between some voiced and voiceless consonants (Eggermont, 1995). Furthermore, a right-ear-advantage could be detected for presenting the gap-stimuli to different ears, analogous to the well known right-ear-advantage in phoneme discrimination (Vroom *et al.,* 1977). Therefore, by using the Gap task as an explicit nonspeech stimulus, we were able to investigate the temporal processes involved in stop-consonant discrimination and directly compare it to activity evoked by CV processing.

Our data clearly demonstrate a common neural substrate for processing Gap and CV tasks as revealed by overlapping activity. Irrespective of the TPA, the comparisons of Gap and Tone task as well as CV and Tone task revealed overlapping areas of activations within the left posterior superior temporal cortex. Additionally, the direct hemispheric comparison also revealed similar overlapping activation pattern. For TPA2 the identical peak of activation for Gap and CV could be observed in the left PT. This observation clearly suggests a general sensitivity of language related areas to rapidly changing acoustic cues rather than a selectivity for speech processing. This notion receives corroboration by previous clinical observations: Patients with acquired brain lesions and aphasia (Efron, 1963; Swisher & Hirsh, 1972), children with general language-learning disabilities (Tallal & Piercy, 1973; Tallal & Stark, 1981) and children and adults with dyslexia (Tallal, 1980) have been shown impaired in the identification of temporal sequences of non-verbal stimuli. Furthermore, it has been shown that children with reading disabilities are deficient in phoneme perception which is reflected by inconsistent labelling of tokens in a Voice Onset Time (VOT) series (Tallal, 1980; Tallal, Miller, & Fitch, 1993). Notably, these children also perform less consistently in labelling of Tone Onset Time tokens (Breier *et al.,* 2001), and exhibit poorer auditory order thresholds (Von Steinbuchel, 1998). Furthermore, it is known that the ability for phoneme discrimination in these children could be increased by a behavioural training using more salient versions of the rapidly changing elements in the acoustic waveform of speech (Merzenich *et al.,* 1996; Tallal *et al.,* 1996).

Notwithstanding our emphasis on rapid temporal information processing we are aware of further important temporal acoustic features in phonetic perception. Here, we identified the neural underpinnings of temporal cues in speech like VOT.

Our present results reasonably contribute to the ongoing debate on hemispheric differences in language processing. Two recent theoretical models on lateralized auditory processing proposed a left auditory cortex superiority for temporal processing and a right hemispheric specialisation for spectral processing (Poeppel, 2003; Zatorre *et al.,* 2002). Although differing in particular points, the two theoretical suggestions are similar in that they argue for a predetermined architecture of the left auditory cortex for analyzing speech-like signals. This functional principle implemented in the left auditory cortex is assumed to allow efficient processing of short acoustic segments while the right auditory cortex is more involved in frequency analysis. In contrast to the notable consensus about the priority of the right hemisphere for processing of spectral, melodic information in auditory sounds, there seems to be conflicting positions about the preference of the left hemisphere for rapid temporal information. According to a recent review paper (Scott & Wise, 2004), several neuroimaging studies comparing slow and fast spectral changes failed to show lateralized temporal lobe responses to rapidly changing sounds (Hall *et al.,* 2002; Johnsrude *et al.,* 1997; Temple *et al.,* 2000). These results made the authors assume that - regardless of simple temporal/spectral dichotomy - the left superior temporal cortex processes any auditory stimulus with sufficient spectro-temporal structural details such as speech. In addition, the right hemisphere is more strongly involved in processing stimuli with dynamic, melodic pitch information.

The present study adds supporting evidence to the theoretical assumptions outlined above. Our data shows an exclusive left hemisphere response to rapid acoustic cues which is in agreement with the view of hemispheric specialization for rapid and slow temporal information (Poeppel, 2003). Since we used nonspeech stimuli that comprise sufficient spectro-temporal details as speech sounds, our data is also in harmony with the

latter proposal (Scott *et al.,* 2004). However, the present data point out, that the similar spectro-temporal structures of speech and nonspeech sounds per se trigger left hemispheric processing and evidences left hemispheric specialisation for rapid temporal information processing independent of linguistic demands.

Conclusions

In this paper we applied a clustered sparse temporal fMRI design to examine the neural underpinning of rapid temporal auditory processing. We have demonstrated that processing rapidly changing cues as required by gap detection exclusively engaged the left auditory cortex. Furthermore, the data suggest that rapid temporal information processing provides the functional basis for phonetic perception.

The data clearly adds weight to recent models on speech perception which propose a differential contribution of left and right brain areas to rapid and slow temporal processing. In sum, the present study supports the assumption of a shared neural network for rapid temporal information processing within the auditory domain for both speech and nonspeech signals.

Acknowledgements

During the preparation of this manuscript Martin Meyer was supported by Schweizer National Fonds (Swiss National Foundation) SNF 46234101. We also would like to thank H.-J. Heinze for providing MR-scanner time.

Abbreviations

EPI, echo planar imaging; FMRI, functional magnetic resonance imaging; TE, echo time; TR, repetition time; HG, Heschl gyrus; PT, planum temporale; STG, superior temporal gyrus;

Reference List

Amaro, E., Williams, S.C., Shergill, S.S., Fu, C.H., MacSweeney, M., Picchioni, M.M., Brammer, M.J. & McGuire, P.K. (2002) Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. J. Magn Reson. Imaging, 16, 497-510.

Annett, M. (1992) Five tests of hand skill. Cortex, 28, 583-600.

Belin, P., Zatorre, R.J. & Ahad, P. (2002) Human temporal-lobe response to vocal sounds. Brain Res. Cogn Brain Res., 13, 17-26.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P. & Pike, B. (2000) Voice-selective areas in human auditory cortex. Nature, 403, 309-312.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.C. & Samson, Y. (1998) Lateralization of speech and auditory temporal processing. J. Cogn Neurosci., 10, 536-540.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N. & Possing, E.T. (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex, 10, 512-528.

Breier, J.I., Gray, L., Fletcher, J.M., Diehl, R.L., Klaas, P., Foorman, B.R. & Molis, M.R. (2001) Perception of voice and tone onset time continua in children with dyslexia with and without attention deficit/hyperactivity disorder. J. Exp. Child Psychol., 80, 245-270.

Brown, S. & Nicholls, M.E. (1997) Hemispheric asymmetries for the temporal resolution of brief auditory stimuli. Percept. Psychophys., 59, 442-447.

Bryden, P. (1982) Laterality: Functional Asymmetry in the Intact Brain Academic Press, Ney York.

Davis, M.H. & Johnsrude, I.S. (2003) Hierarchical processing in spoken language comprehension. J. Neurosci., 23, 3423-3431.

Efron, R. (1963) Temporal Perception, Aphasia and D'ej'a vu. Brain, 86, 403-424.

Eggermont, J.J. (1995) Neural correlates of gap detection and auditory fusion in cat auditory cortex. Neuroreport, 6, 1645-1648.

Emmerich, D.S., Pitchford, L.J., Joyce, C.M. & Koppell, S.M. (1981) Laterality effects in response to offsets of tonal stimuli. Neuropsychologia, 19, 227-234.

Evans, A.C., Collins, D.L., Mills, S.R., Brown, E.D., Kelly, R.L. & Peters, T.M. (1993) 3D statistical neuroanatomical models from 305 MRI volumes. Proc. IEEE Nucl. Sci. Symp. Med. Imaging, 3, 1813-1817.

Fernandez, G., de Greiff, A., von Oertzen, J., Reuber, M., Lun, S., Klaver, P., Ruhlmann, J., Reul, J. & Elger, C.E. (2001) Language mapping in less than 15 minutes: real-time functional MRI during routine clinical investigation. Neuroimage., 14, 585-594.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S. & Turner, R. (1995) Analysis of fMRI time-series revisited. Neuroimage., 2, 45-53.

Gaab, N., Gaser, C., Zaehle, T., Jancke, L. & Schlaug, G. (2003) Functional anatomy of pitch memory--an fMRI study with sparse temporal sampling. Neuroimage., 19, 1417-1426.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M. & Bowtell, R.W. (1999) "Sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp., 7, 213-223.

Hall, D.A., Johnsrude, I.S., Haggard, M.P., Palmer, A.R., Akeroyd, M.A. & Summerfield, A.Q. (2002) Spectral and temporal processing in human auditory cortex. Cereb. Cortex, 12, 140-149.

Jancke, L., Wustenberg, T., Scheich, H. & Heinze, H.J. (2002) Phonetic perception and the temporal cortex. Neuroimage., 15, 733-746.

Joanisse, M.F. & Gati, J.S. (2003) Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. Neuroimage., 19, 64-79.

Johnsrude, I.S., Zatorre, R.J., Milner, B.A. & Evans, A.C. (1997) Left-hemisphere specialization for the processing of acoustic transients. Neuroreport, 8, 1761-1765.

Kimura, D. (1961) Some effects of temporal-lobe damage on auditory perception. Can.J.Psychol., 15, 156-165.

Kimura, D. & Folb, S. (1968) Neural processing of backwards-speech sounds. Science, 161, 395-396.

Liberman, A.M. & Mattingly, I.G. (1985) The motor theory of speech perception revised. Cognition, 21, 1-36.

Liberman, A.M. & Whalen, D.H. (2000) On the relation of speech to language. Trends Cogn Sci., 4, 187-196.

Liegeois-Chauvel, C., de Graaf, J.B., Laguitton, V. & Chauvel, P. (1999) Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb. Cortex, 9, 484-496.

Luders, E., Steinmetz, H. & Jancke, L. (2002) Brain size and grey matter volume in the healthy human brain. Neuroreport, 13, 2371-2374.

Merzenich, M.M., Jenkins, W.M., Johnston, P., Schreiner, C., Miller, S.L. & Tallal, P. (1996) Temporal processing deficits of language-learning impaired children ameliorated by training. Science, 271, 77-81.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G. & von Cramon, D.Y. (2002) FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum. Brain Mapp., 17, 73-88.

Meyer, M., Friederici, A.D. & von Cramon, D.Y. (2000) Neurocognition of auditory sentence comprehension: event related fMRI reveals sensitivity to syntactic violations and task demands. Brain Res. Cogn Brain Res., 9, 19-33.

Mills, L. & Rollman, G.B. (1979) Left hemisphere selectivity for processing duration in normal subjects. Brain Lang, 7, 320-335.

Mills, L. & Rollman, G.B. (1980) Hemispheric asymmetry for auditory perception of temporal order. Neuropsychologia, 18, 41-48.

Phillips, D.P., Hall, S.E., Harrington, I.A. & Taylor, T.L. (1998) "Central" auditory gap detection: a spatial case. J. Acoust. Soc. Am., 103, 2064-2068.

Phillips, D.P., Taylor, T.L., Hall, S.E., Carr, M.M. & Mossop, J.E. (1997) Detection of silent intervals between noises activating different perceptual channels: some properties of "central" auditory gap detection. J. Acoust .Soc. Am., 101, 3694-3705.

Poeppel, D. (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication, 41, 245-255.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H.J. & Zilles, K. (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage, 13, 669-683.

Robin, D.A., Tranel, D. & Damasio, H. (1990) Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. Brain Lang, 39, 539-555.

Schwartz, J. & Tallal, P. (1980) Rate of acoustic change may underlie hemispheric specialization for speech perception. Science, 207, 1380-1381.

Scott, S.K. & Johnsrude, I.S. (2003) The neuroanatomical and functional organization of speech perception. Trends Neurosci., 26, 100-107.

Scott, S.K. & Wise, R.J. (2004) The functional neuroanatomy of prelexical processing in speech perception. Cognition, 92, 13-45.

Swisher, L. & Hirsh, I.J. (1972) Brain damage and the ordering of two temporally successive stimuli. Neuropsychologia, 10, 137-152.

Talairach, J. & Tournoux, P. (1988) Co-palanar Stereotaxis Atlas of the Human Brain Thieme, New York.

Tallal, P. & Piercy, M. (1973) Developmental aphasia: impaired rate of non-verbal processing as a function of sensory modality. Neuropsychologia, 11, 389-398.

Tallal, P. & Newcombe F. (1978) Impairment of auditory perception and language comprehension in dysphasia. Brain Lang, 5, 13-34.

Tallal, P. (1980) Auditory temporal perception, phonics, and reading disabilities in children. Brain Lang, 9, 182-198.

Tallal, P. & Stark, R.E. (1981) Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. J. Acoust. Soc. Am., 69, 568-574.

Tallal, P., Miller, S. & Fitch, R.H. (1993) Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann. N.Y. Acad. Sci., 682, 27-47.

Tallal, P., Miller, S.L., Bedi, G., Byma, G., Wang, X., Nagarajan, S.S., Schreiner, C., Jenkins, W.M. & Merzenich, M.M. (1996) Language comprehension in language-learning impaired children improved with acoustically modified speech. Science, 271, 81-84.

Temple, E., Poldrack, R.A., Protopapas, A., Nagarajan, S., Salz, T., Tallal, P., Merzenich, M.M. & Gabrieli, J.D. (2000) Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. Proc. Natl. Acad. Sci. U.S.A, 97, 13907-13912.

Von Steinbuchel, N. & Poppel, E. (1993) Domains of rehabilitation: a theoretical perspective. Behav. Brain Res., 56, 1-10.

Von Steinbuchel, N. (1998) Temporal ranges of central nervous processing: clinical evidence. Exp. Brain Res., 123, 220-233.

Von Steinbuchel, N., Wittmann,M., Strasburger,H. & Szelag,E. (1999) Auditory temporal-order judgement is impaired in patients with cortical lesions in posterior regions of the left hemisphere. Neurosci. Lett., 264, 168-171.

Vroom, P.A., Timmers, H. & Tempelaars, S. (1977) On the hemispheric representation of time. In Dornic, S. (ed), Attention and Performance VI. Erlbaum, Hillsdale, New Jersey, pp. 231-245.

Westbury, C.F., Zatorre, R.J. & Evans, A.C. (1999) Quantifying variability in the planum temporale: a probability map. Cereb. Cortex, 9, 392-405.

Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K. & Warburton, E.A. (2001) Separate neural subsystems within 'Wernicke's area'. Brain, 124, 83-95.

Zatorre, R.J., Belin, P. & Penhune, V.B. (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci., 6, 37-46.

4. Investigations

The following section presents four experiments (three fMRI and one EEG study) investigating the nature of temporal and spectral processing in the auditory domain.

The first fMRI experiment entitled "*Spectral-temporal processing during speech perception involves left posterior auditory cortex*" investigated the neural underpinnings of spectral-temporal integration during speech perception. The study consisted of a behavioral pilot study and a follow-up fMRI-experiment to elucidate the neural underpinnings of basic linguistic processing. We presented participants with sine-wave analogues that could be either perceived as non-speech auditory forms (naïve condition) or after instruction and short practice as speech (informed condition). Behavioral results revealed a difference in the processing mode; spectral-temporal integration occurred during speech perception, but not when stimuli were perceived as non-speech. In terms of neuroimaging we observed an activation increase in left posterior primary and secondary auditory cortex, namely Heschl's gyrus and planum temporale encroaching onto the superior temporal sulcus reflecting a shift from auditory to speech perception. This finding demonstrates that the left posterior superior temporal lobe is essential for spectral-temporal processing during speech perception.

The second fMRI experiment entitled "*The neural correlate of speech rhythm as evidenced by meter processing: an fmri study*" investigates the neural correlates of rhythm processing in speech perception using a comparison of German pseudo-sentences spoken with an exaggerated (metric) or a conversational (non-metric) rhythm in the context of an auditory clustered-temporal-sampling fMRI experiment. For the perception of speech rhythm, suprasegmental cues such as syllable duration, syllable stress or pause are most important. Therefore, the analysis of these acoustic suprasegmental speech characteristics mainly relies on the perception of

temporal intervals and thus the perception of speech rhythm incorporates an appropriate tool to investigate auditory acoustic temporal processing beyond the perception on the level of segmental speech characteristics such as consonants and vowels. Results of the study indicate a function of SMA and insula beyond motor timing and speak for a role of these brain areas in the perception of temporal intervals. Furthermore, the data demonstrate a specific task-related function of the right IFG in the processing of accent patterns and show right secondary auditory cortex involvement in the explicit perception of auditory cues and that activity in the right secondary auditory cortex can be modulated by top-down mechanisms.

The third fMRI experiment entitled "*Evidence for a general segmental processing device in the human auditory dorsal stream*" investigates the functional organization of sub-lexical auditory perception with specific respect to auditory spectro-temporal processing in speech and non-speech sounds. Participants were presented with verbal and nonverbal auditory stimuli with systematic variations of the spectral or temporal acoustic characteristics in the context of a sparse event-related functional magnetic resonance imaging (fMRI) study. Results of this investigation show a subdominant left hemisphere involvement during the processing of rapidly changing temporal characteristics regardless of the speechness of the sounds. In particular, when participants had to attend to subtle temporal modulations within speech and non-speech stimuli we observed significant responses in the parietal operculum and the frontal operculum on the left hemisphere. In contrast, we revealed no lateralization effects for spectral processing. When participants had to listen to changes in the frequency content of speech and non-speech stimuli, activations of the medial temporal gyrus und superior temporal sulcus bilaterally were explored.

In the fourth EEG experiment entitled "Electrical brain imaging reveals overlapping activity elicited by temporal cue processing in speech and non-speech sounds" scalp AEPs in response to CV-syllables and non-speech analogues with varying VOT and noise-onset-time (NOT), respectively were

recorded. This study investigates the neural coding of acoustic characteristics underlying speech perception. Results showed that the characteristic AEP waveform in response to consonant-vowel-syllables can be resembled by non-speech sounds with similar temporal characteristics, and that the primary auditory cortex is asymmetrically activated in favoring the left hemisphere independently of the stimulation used.

4.1 Study D:

Spectral-temporal processing during speech perception involves left posterior auditory cortex

MARTIN MEYER[1], **TINO ZAEHLE**[1], VIKTORIA-ELENI GOUNTOUNA[2],

ANTHONY BARRON[3], LUTZ JANCKE[1], ALICE TURK[3]

[1]DEPARTMENT OF NEUROPSYCHOLOGY, INSTITUTE FOR PSYCHOLOGY, UNIVERSITY OF ZURICH, SWITZERLAND
[2]DIVISION OF PSYCHIATRY, UNIVERSITY OF EDINBURGH, UNITED KINGDOM
[3]DEPARTMENT OF THEORETICAL AND APPLIED LINGUISTICS, UNIVERSITY OF EDINBURGH, UNITED KINGDOM

Published in:

Neuroreport, 16: 1985-1989, 2005.

Abstract

This fMRI-study investigates the neural underpinnings of spectral-temporal integration during speech perception. Participants performed an auditory discrimination task on a set of sine wave analogues that could be either perceived as non-speech or speech. Behavioural results revealed a difference in the processing mode; spectral-temporal integration occurred during speech perception, but not when stimuli were perceived as non-speech. In terms of neuroimaging we observed an activation increase in left posterior primary and secondary auditory cortex, namely Heschl's gyrus and planum temporale encroaching onto the superior temporal sulcus reflecting a shift from auditory to speech perception. This finding demonstrates that the left posterior superior temporal lobe is essential for spectral-temporal processing during speech perception.

Introduction

A predominant role of the left hemisphere in speech perception has been observed from the very outset of cognitive neuropsychology (Josse and Tzourio-Mazoyer, 2004). However, the existence of left hemisphere speech-specifc regions has been significantly challenged. One approach to reconcile the notion of a left hemisphere preference for speech and a position which denies the existence of a neurally confined speech centre advocates the idea of general left hemisphere responsitivity to temporal information available in speech signals and to its role in integrating these cues with other spectral or spectral-temporal cues. These cues are not exclusive to speech, but successful speech identification requires attention to these cues, as well as their successful integration (Tallal *et al.,* 1993; Tallal and Newcombe, 1978).

Hence, recent neurobiological models of speech perception have been incorporating the role of temporal information processing during speech perception (Hickok and Poeppel, 2004; Poeppel, 2003; Zatorre and Belin, 2001). These approaches are principally parallel in that they stipulate that the left auditory cortex specializes in processing stimuli requiring rapid temporal resolution.

Recent brain imaging studies have been providing corroborating evidence insofar as they demonstrate a preference of left auditory cortices for rapidly changing acoustic cues available in speech and nonspeech sounds (Boemio *et al.,* 2005; Jancke *et al.,* 2002; Joanisse and Gati, 2003; Zaehle *et al.,* 2004). These findings validate the existence of an auditory processing device localized in the left posterior auditory cortex which preferentially operates on quick temporal acoustic changes and is thus indirectly most amenable to the acoustic composition of human speech. In particular, the planum temporale (PT) which is an associative auditory region posterior to Heschl's gyrus (HG) has been coined a cortical 'computational hub' that mainly serves spectral-temporal integration of the type required to process rapidly changing cues among other things during auditory perception (Griffiths and Warren, 2002). Furthermore, recent brain imaging studies have compellingly demonstrated that the left PT does not holistically mediate

speech perception *per se,* but principally governs the analysis of rapidly changing acoustic cues available in both speech and nonspeech stimuli (Jancke *et al.,* 2002; Zaehle *et al.,* 2004). Thus, these findings qualify the view that early analyses of the physical attributes of speech signals occur in bilateral posterior superior temporal gyrus (STG) and PT (Boemio *et al.,* 2005; Hickok and Poeppel, 2004) with the left posterior auditory cortex being specifically adept at analyzing temporal arrangement of spectral-temporal features, e.g. formant transitions and voice-onset-time (VOT). However, this holds principally for all short segmental information as universal temporal sensitivity must hold across speech and nonspeech sounds. The present study consisted of a behavioral pilot study and a follow-up fMRI-experiment which we set out to further elucidate the neural underpinnings of basic linguistic processing. We presented participants with sine-wave analogues that could be either perceived as non-speech auditory forms (naïve condition) or after instruction and short practice as speech (informed condition) (Best *et al.,* 1981). We propose that analysis of natural speech forms particularly requires decoding the rapidly changing fine-grained temporal information and the integration of temporal information with other spectral cues which is supposed to preferentially drive the left posterior auditory cortex.

Methods

Stimuli and task

Stimuli were noise plus sine–wave analogues of the spoken words *say* and *stay*. They were synthesised from the recordings of a male native speaker. The records were used to extract fricative noise frequency distributions information and the first three formant trajectories of the vowel, to estimate of the fricatives, stop closure, and vowel segment durations. After determination of endpoints two continua were generated: one continuum has vowel F1 onset (496 Hz) of the word *say* and the other continuum of the word *stay* (449 Hz) with the stop gap duration in both continua varying from 0 to 99 ms in 11 ms steps. Stimuli were evaluated by both naïve and informed listeners. The sounds were mainly described by

naïve listeners as unnatural synthetic sounds, while informed listeners had no difficulty in recognising them as speech. In the behavioural and in the fMRI-study participants performed a two-alternative forced choice AXB identification task. Each event (triple) consisted of three sounds presented in rapid succession, the first sound always was the most extreme version of *say* and the third sound always the most extreme version of *stay*. The second one was chosen from a reduced set of 12 stimuli, comprising the two continua varying from 11 to 66 ms. The order of presentation of the changing stimulus was pseudo-randomised and a varying interval of 5 to 7 seconds between the triplets was used. Eight experimental periods, i.e. repetitions of the whole set, were presented. Periods were separated by rest periods without visual or auditory stimulation. Each rest period lasted ten seconds. In the fMRI study the participants performed the task twice. After completing the task in the naive state and while they were still lying in the scanner they were informed that the sounds they were hearing can be perceived as speech. They then practiced again with modified instructions and new labels for the stimuli before performing the same task in the informed state. In the naive condition the three sounds of the triplet were labelled tokens 1, 2, and 3 and participants had to indicate whether token 2 was more like token 1 or like token 3. In the informed condition they had to indicate whether the second word was more like *say* or *stay*.

Participants

In a pilot study 26 healthy, right-handed native speakers of English performed the task in the naive condition while being placed in an fMRI simulator. At the beginning participants did two short practice runs. An accuracy score of 90% in both was used as criterion for participation. Based on analysis of behavioural performance we excluded 11 subjects from further analysis as they turned out unable to discern any differences between the endpoints, started to hear the stimuli as speech, or explored shifting strategies. The remaining fifteen volunteers were judged as appropriate for participating in the fMRI experiment but only thirteen subjects finally partook in the fMRI study. Subjects gave written informed

consent for the study according to a protocol approved by the local Lothian NHS Ethics Committee.

fMRI scanning

Functional MR imaging was performed on a 1.5 Tesla whole-body MRI system (GE Medical Systems), equipped with echo planar imaging (EPI) capability and standard circular polarized head coil. Using a mid-sagittal scout image, 14 axial slices (slice thickness 5 mm; no inter slice gap) of a T2* weighted gradient echo EPI sequence were acquired parallel to the bicommissural plane (TR: 2 s, TE: 40 ms, voxel size: 3.75 x 3.75 x 5 mm^3). Total scanning time per block was 1060 seconds, during which 508 volumes were collected.

Data Analysis

Realignment, spatial normalization, spatial smoothing (10 mm FWHM) and image analysis was performed using MATLAB 6.5 (Mathworks) and the SPM99 software package. Condition and subject effects were estimated using the GLM (Friston *et al.,* 1995). After estimation of model parameters for each subject, an analysis of variance was calculated for the whole group, using the individual contrast images of all auditory stimulation conditions. To test hypotheses about regionally specific condition effects, linear contrasts were employed in the context of a random effects procedure. The resulting SPM{T} for this effect were thresholded at a T-value of 3.58 (P = 0.001, uncorrected for multiple comparisons). Subsequently anatomical ROIs were defined based on anatomical definitions of functional areas including: HG, PT, planum polare (PP), and STS in both hemispheres. For each ROI the mean T-values from each subject and each condition (uniformed and informed) were subjected to a two-factor repeated-measure ANOVA. Post hoc analyses were carried out using non-parametric Wilcoxon-test for related samples.

Results

Behavioural data

Based on the behavioural data, 8 individuals (5 male, age range 18–24, mean 20.4) out of 13 persons who initially partook in the fMRI experiment entered data analysis. We excluded 5 subjects as they failed to perceive the stimuli as speech in the informed condition or explored shifting strategies. Figure 1 shows mean performance data separated for the "uninformed" and "informed condition". Horizontal response curves for uninformed subjects suggest that they did not rely on the changing duration of the silent interval in identifying presented stimuli. Separated curves indicate that they based their identification solely on the dynamic spectral cues. In the informed condition, participants showed more *stay* responses to longer silent intervals as well as more *stay* responses to lower F1 onset frequencies, as evidenced by their rising, separated response curves. This pattern of results suggests successful integration of both, temporal and spectral cues.

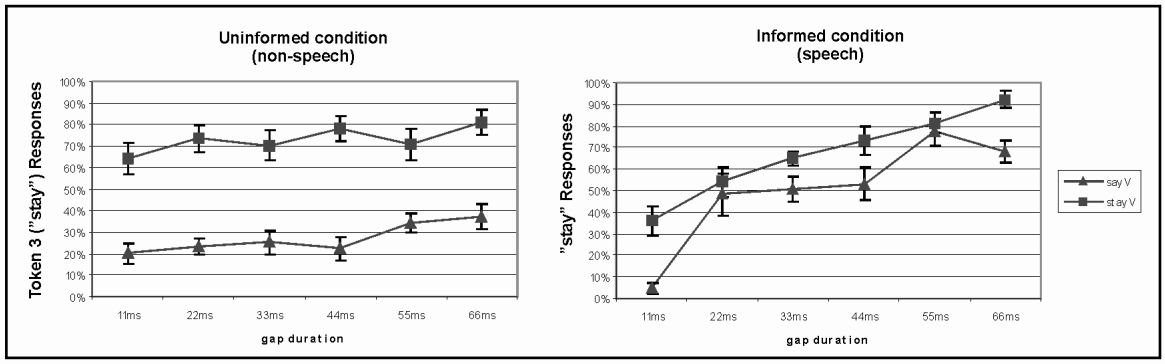

Figure 1. Mean performance data (n=8) separated for the uninformed "non-speech" condition (left) and informed "speech" condition (right). The figures shows mean Token 3 and 'stay' responses (percent) respectively graphed against gap duration. The graph on the left demonstrates that in the uninformed condition participants were using spectral information only to classify the stimuli and did not pay any attention to the varying stop gap durations. Data collected during "informed condition" indicate that participants integrated both spectral and temporal information when listening to speech.

Functional MR Imaging

We compared hemodynamic responses to all sine wave replicas with responses to zero trials to identify areas indicative of auditory stimulation. Generally, the random-effects analysis revealed that listening to stimuli

involved bilateral supratemporal regions including HG, PT, PP, and portions along the dorsal superior temporal sulcus (STS).

The ROI analyses revealed significant main effects “mode” (informed vs. uninformed) in HG ($F(1,7)= 6.74$ $P < 0.05$) and PT ($F(1,7)=17.243$ $P < 0.005$) and a trend in STS ($F(1,7)=3.860$, $P= 0.09$). These results show changes in local activity reflecting the perceptual switch (“effect of speechness”). We also found a significant main effect “hemisphere” (left vs. right) for STS ($F(1,7)=17.478$, $P = 0.004$). Finally, we observed a significant interaction “condition” x “hemisphere” in HG ($F(1,7)= 11.214$ $P = 0.012$) which demonstrates a special sensitivity of left HG to speech stimuli. No additional main effect or interaction was significant.

Post-hoc-tests validate the aforementioned results. Figure 2 depicts that the “informed condition” involved left HG (P=0.017), left PT (P=0.017), and left STS (P=0.035) significantly stronger than the “naïve” mode.

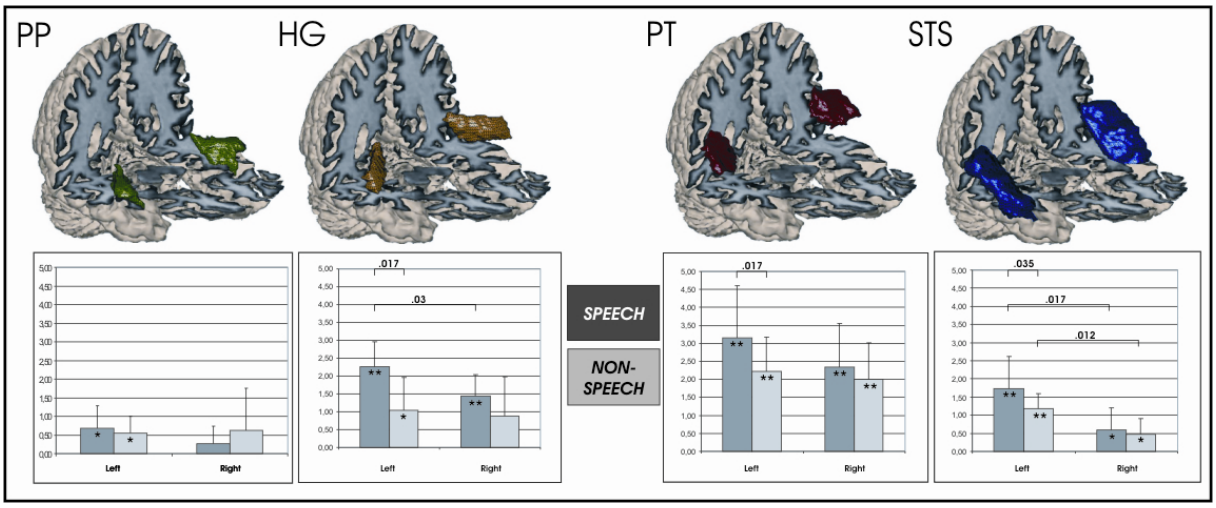

Figure 2. ROI analysis. The upper panel illustrates the position and size of four bilateral temporal ROIs (PP=planum polare, HG=Heschl’s gyrus, PT=planum temporale, STS=superior temporal sulcus). The lower panel depicts mean and standard error of the effect sizes (mean t-values) collected from these ROIs. Asterisks indicate significance level (*P=0.05, **P=0.001) as revealed by one-sample t-test. The HG, PT and STS of the left hemisphere exhibit a significant activational increase reflecting the switch from non-speech perception (uninformed condition) to speech perception (informed condition). For the informed condition during which subjects perceived the stimulation as speech, HG and STS explored significantly stronger involvement of the LH directly compared to the contralateral area.

Furthermore, the speech perception mode (“informed condition”) corresponded to a significant functional leftward bias in HG (P=0.03) and

STS (P=0.017). With regard to STS we noted the same starker recruitment of the left hemisphere for nonspeech perception ("naïve mode").

Discussion

We set out the present study to provide evidence that basic linguistic processing is contingent upon successful temporal processing and spectral-temporal integration. Our hypothesis derives from the view that analysis of brief temporal auditory cues (e.g. VOT, formant transitions) is an important part of speech perception (Tallal *et al.,* 1993; Tallal and Newcombe, 1978). Indirect support for the eminent relevance of temporal information for proper speech perception is evident from a study which demonstrated sufficient comprehension of degraded speech even after partial removal of spectral information (Shannon *et al.,* 1995). Further supporting evidence comes from cases of language-learning impairment where children often reveal auditory perception deficits for brief temporal auditory features (Tallal *et al.,* 1996). Our behavioral data clearly support this view as we demonstrate a perceptual switch from reliance on a single cue (spectral information) in the naïve condition to spectral-temporal cue integration in the informed condition as soon as participants experienced the sine wave analogues as speech. Thus, our data solidly emphasize the pivotal role of auditory temporal processing for proper speech identification.

Secondly, we expected to add further support to the notion of left-hemisphere preference for speech since former studies have demonstrated the left posterior auditory cortex as specifically governing acoustic processes that are critical to the analysis of spoken language (Josse and Tzourio-Mazoyer, 2004). Indeed, we are capable of buttressing the increasing body of research which postulates that left auditory cortex is preferentially driven by briefly changing temporal acoustic transitions which underlie phonological processing rather than being selectively triggered by speech *per se* (Jancke *et al.,* 2002; Liegeois-Chauvel *et al.,* 1999; Oram Cardy *et al.,* 2005; Poeppel, 2003; Zaehle *et al.,* 2004). To elaborate, the current study identified a signal increase corresponding to the switch from non-speech to speech perception

in three distinct supratemporal areas (HG, STS, PT) which we discuss in turn.

Converging data from neurophysiological and imaging studies emphasize the responsiveness of the left primary (Belin *et al.,* 1998; Liegeois-Chauvel *et al.,* 1999; Trebuchon-Da *et al.,* 2005) and secondary (Jancke *et al.,* 2002; Joanisse and Gati, 2003) auditory fields to rapid acoustic transitions in speech and nonspeech stimuli. Interestingly, in the informed condition the left HG turned out significantly stronger activated than the contralateral hemisphere which suggests that neural ensembles which are specifically equipped to detect temporal patterns typical for speech also reside in the left HG.

Notably, activation in the STS forms the same pattern which also points to a sensitivity of left STS to auditory perception. Noteworthy, the STS has recently been tied to serve auditory functions (Warren *et al.,* 2005). Furthermore, fMRI based evidence has been specifying the role of STS since this region sensitively responds to phonetic cues (Jancke *et al.,* 2002; Rimol *et al.,* 2005) and even appears to be preferentially driven by rapidly changing acoustic cues (Boemio *et al.,* 2005) which is in keeping with our postulation. Additionally, one recent fMRI-study revealed a similar left-sided circuitry including posterior parts of the left STS which is amenable to top-down processing of ambiguous speech/nonspeech stimuli (Dhaene-Lambertz *et al.,* 2005). The authors' conclusion slightly differs from our interpretation in that they more generally stipulate the left posterior STS as susceptible to syllable processing. Surprisingly, another recent fMRI-study also investigating the neural substrates of perceptual switch from naïve to informed processing did not find an increase, but a signal decrease in left posterior STG in the informed condition (Liebenthal *et al.,* 2003). The authors speculate that an adaptation of the general auditory system resulting from lifelong exposure to phonetic patterns accounts for this finding. However, this observed "decrease" is at odds with the present models of speech perception (Hickok and Poeppel, 2004) and hence lacks plausibility.

A slightly different picture emerges from observation of PT. The involvement of right PT is not significantly weaker relative to the left PT which provides support in favor of an important role of both left and right PT during speech perception. As proper speech perception is reliant on the integration of both temporal and spectral information our findings concurs with the current opinion of the bilateral PT as a 'computational hub' which is not considered a speech decoding area per se but primarily serves spectral-temporal integration (Boemio *et al.,* 2005; Griffiths and Warren, 2002). However, our study cannot definitely clarify whether the PT serves as spectral-temporal integrator or as simple temporal processor.

Principally, a present model assumes that left auditory cortex is more proficient at processing stimuli requiring enhanced temporal resolution whilst the right auditory cortex is better adept at processing spectral information (Zatorre and Belin, 2001). A complimentary model also proposes a leftward lateralization for analysis of short segments which can be considered the most essential acoustic units for phonetic perception (Poeppel, 2003). Our study is principally in harmony with the two complementary models and even specifies them by describing in how far distinct regions of the human auditory region demonstrate differential involvement even pertaining to hemispheric lateralization.

Our conclusion that the left auditory cortex is better equipped to process rapid temporal events is also clearly bolstered by anatomical observations which indicate that columnar and connectional architecture within left auditory fields is specifically proficient at extracting crucial temporal information from the incoming stream of auditory inflow (Hutsler and Galuske, 2003).

Conclusion

We performed this study to elucidate the neural substrates of spectro-temporal processing during speech perception by presenting participants sinusoids which could be either perceived as speech or non-speech. Perceptually switching from spectral processing (only) to an integrative

spectral-temporal processing in the auditory domain (during auditory speech processing) brought on a signal increase in left temporal regions including HG, PT, and STS which may reflect an enhanced analysis of temporal acoustic cues, and their integration with spectral information. Thus, our study lends further credence to the view that computation of temporal acoustic events is an indispensable part of auditory speech perception. Furthermore we show that the left auditory cortex is specifically proficient at decoding briefly temporal changing temporal cues which may explain the left hemisphere preference for speech.

Acknowledgements:

During the preparation of this manuscript M. M. was supported by Schweizer National Fonds (Swiss National Foundation) SNF 46234101 and T. Z. was supported by SNF 46234103. We also would like to thank the Developmental Trust Research Fund (DTRF), University of Edinburgh, for sponsoring MR-scanner time and T. Wustenberg for preparing figures.

Reference List

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.C., Samson, Y., 1998. Lateralization of speech and auditory temporal processing. J.Cogn Neurosci. 10, 536-540.

Best, C.T., Morrongiello, B., Robson, R., 1981. Perceptual equivalence of acoustic cues in speech and nonspeech perception. Percept.Psychophys. 29, 191-211.

Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat.Neurosci. 8, 389-395.

Dhaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., Dehaene, S., 2005. Neural correlates of switching from auditory to speech perception. Neuroimage 24, 21-33.

Friston, K.J., Holmes, A.P., Worsley, K.P., Poline, J.B., Frith, C.D., Frackowiak, R.S., 1995. Statistical parameter maps in functional imaging: A general linear approach. Hum.Brain Mapp. 2, 189-210.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348-353.

Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67-99.

Hutsler, J., Galuske, R.A., 2003. Hemispheric asymmetries in cerebral cortical networks. Trends Neurosci. 26, 429-435.

Jancke, L., Wustenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. Neuroimage. 15, 733-746.

Joanisse, M.F., Gati, J.S., 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. Neuroimage. 19, 64-79.

Josse, G., Tzourio-Mazoyer, N., 2004. Hemispheric specialization for language. Brain Res.Brain Res.Rev. 44, 1-12.

Liebenthal, E., Ellingson, M.L., Spanaki, M.V., Prieto, T.E., Ropella, K.M., Binder, J.R., 2003. Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. Neuroimage. 19, 1395-1404.

Liegeois-Chauvel, C., de Graaf, J.B., Laguitton, V., Chauvel, P., 1999. Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb.Cortex 9, 484-496.

Oram Cardy, J.E., Flagg, E.J., Roberts, W., Brian, J., Roberts, T.P., 2005. Magnetoencephalography identifies rapid temporal processing deficit in autism and language impairment. Neuroreport 16, 329-332.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication 41, 245-255.

Rimol, L.M., Specht, K., Weis, S., Savoy, R., Hugdahl, K., 2005. Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. Neuroimage 26, 1059-1067.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303-304.

Tallal, P., Miller, S., Fitch, R.H., 1993. Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann.N.Y.Acad.Sci. 682, 27-47.

Tallal, P., Miller, S.L., Bedi, G., Byma, G., Wang, X., Nagarajan, S.S., Schreiner, C., Jenkins, W.M., Merzenich, M.M., 1996. Language comprehension in language-learning impaired children improved with acoustically modified speech. Science 271, 81-84.

Tallal, P., Newcombe, F., 1978. Impairment of auditory perception and language comprehension in dysphasia. Brain Lang 5, 13-34.

Trebuchon-Da, F.A., Giraud, K., Badier, J.M., Chauvel, P., Liegeois-Chauvel, C., 2005. Hemispheric lateralization of voice onset time (VOT) comparison between depth and scalp EEG recordings. Neuroimage 27, 1-14.

Warren, J.D., Jennings, A.R., Griffiths, T.D., 2005. Analysis of the spectral envelope of sounds by the human brain. Neuroimage 24, 1052-1057.

Zaehle, T., Wustenberg, T., Meyer, M., Jancke, L., 2004. Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur.J.Neurosci. 20, 2447-2456.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb.Cortex 11, 946-953.

4.2 Study E:

The neural correlate of speech rhythm as evidenced by meter processing: an fmri study.

EVELINE GEISER[1], **TINO ZAEHLE**[1], LUTZ JANCKE[1] & MARTIN MEYER[2]

[1]Departement of Neuropsychology, University of Zurich, Switzerland
[2]Institute of Neuroradiology, University Hospital of Zurich, Switzerland

Accepted for publication in:

Journal of Cognitive Neuroscience

Abstract

The present study investigates the neural correlates of rhythm processing in speech perception. German pseudo-sentences spoken with an exaggerated (metric) or a conversational (non-metric) rhythm were compared in an auditory fMRI experiment. The subjects had to perform either a rhythm task (explicit rhythm processing) or a prosody task (implicit rhythm processing). The study revealed bilateral activation in the supplementary motor area (SMA) extending into the cingulate gyrus and in the insulae extending into the right basal ganglia (neostriatum, NS), as well as activity in the right inferior frontal gyrus (IFG) related to the performance of the rhythm task. A direct contrast between metric and the non-metric sentences revealed differences in lateralization of activation for metric processing as a function of the explicit and implicit task. Explicit processing revealed activation in the right posterior superior temporal gyrus (pSTG), the right supramarginal gyrus (SMG) and the right parietal operculum (PaOp). Implicit processing showed activation in the left SMG, the left posterior STG, and the left PaOp. The present results indicate a function of SMA and insula beyond motor timing and speak for a role of these brain areas in the perception of temporal intervals. Secondly, the data speak for a specific task-related function of the right IFG in the processing of accent patterns. Finally, the data sustain the assumption that the right secondary auditory cortex is involved in the explicit perception of auditory cues and moreover, that activity in the right secondary auditory cortex can be modulated by top-down processing mechanisms.

Introduction

The flow of speech is commonly perceived as having a particular rhythmic pattern. This rhythm is a pivotal structuring element of speech which is crucially involved in segmentation processes subserving language comprehension and even acquisition.

Recent evidence for this crucial function of speech rhythm has been found in electrophysiological research. Speech segmentation or the parsing of a sentence into words has been shown to rely on global patterns of prosodic phrasing which includes rhythmic and durational grouping as well as tonal pitch (Frazier *et al.,* 2006). At the word level, research has demonstrated that rhythmic patterning of German utterances facilitate spoken word identification (Friedrich *et al.,* 2004b; Steinhauer *et al.,* 1999a). Depending on the rhythmic organization of one's native language different segmentation strategies are observed. For French-speaking adults the syllable appears to be the unit of segmentation (Segui *et al.,* 1990), whereas English-speaking adults are guided by information about typical word-stress patterns (Cutler and Norris, 1988). With respect to the segmentation strategy, the German language is comparable to the English language. The view that the rhythmic properties of language shape listeners' speech processing strategies has led to hypotheses of how infants develop efficient speech segmentation procedures in the acquisition of their mother tongue. There is evidence that infants' sensitivity to linguistic rhythm allows them to discriminate (non-native) languages from different rhythmic classes (Nazzi and Ramus, 2003a) suggesting a pivotal role of rhythm in language acquisition. Similar results have even been reported from primate studies (Tincoff *et al.,* 2005).

The intuitive notion that spoken languages have characteristic underlying rhythmic patterns has resulted in a considerable amount of linguistic research aimed at assessing the exact acoustic characteristics of speech rhythm. This research was driven by the hypothesis of isochrony as an underlying principle of linguistic speech rhythm. Thus, a distinction is

often made between stress-timed languages in which periodicity is assumed to be based on the regular occurrence of stressed syllables (e.g. English, German, Dutch) and syllable-timed languages such as French, Italian, or Spanish in which regularity is taken to be the outcome of all syllables being roughly of the same length (Pike, 1945). However, the phonological evidence for isochrony as the organizational principle of linguistic rhythm is still not unequivocal. As has been shown, the duration of interstress intervals in English is proportional to the number of syllables they contain, but can vary in duration influenced by the specific types of syllables they are comprised of as well as by the position of the interval within the utterance (Bolinger, 1965; Lehiste, 1977). Moreover, new statistical analyses of the speech signal, for instance of consonantal and vocalic intervals (Ramus, Nespor, & Mehler, 1999) or of the auditory prominence of phonetic events (Lee and Todd, 2004) have also been inconclusive with respect to the identification of the acoustic characteristics of speech rhythm. Rhythmic differences between languages, context dependent variability of speech rhythm and inter-individual differences make identifying the acoustic as well as the phonological elements constituting speech rhythm experimentally difficult. Whereas linguists classify speech rhythm together with speech melody as prosodic aspects of speech using the term unspecifically for suprasegmental speech characteristics extending over more than one sound segment or phoneme, psycholinguistic researchers investigate speech rhythm as a concept in its own right. Such studies have included the field of language acquisition (Nazzi and Ramus, 2003b), subsequent language learning (Curtin, 2006), speech segmentation (McQueen *et al.,* 2001) and speech typologies (Ramus *et al.,* 1999).

With respect to the neural correlate of speech rhythm reliable evidence is still sparse. In the context of speech rhythm perception, suprasegmental cues such as syllable duration, syllable stress, or pause are most important, so that it comes as no surprise that neuropsychological research has focussed initially on these phonetic cues. In particular, the occurrence of specific event-related brain–potentials, namely the “closure positive shift”

(CPS) have been associated with the processing of speech accent or speech pauses (Friedrich *et al.,* 2004a; Steinhauer *et al.,* 1999b). However, to our knowledge the perception of speech rhythm as a composition of these elements and expanding over a longer period of time such as the duration of a sentence has never been the subject of investigation in functional brain research.

From the point of view of neuropsychological research, speech rhythm is of particular interest related to the question of functional hemispheric lateralization. Even though the preponderance of the left hemisphere for speech processing is one of the best-evidenced knowledge in cognitive neuroscience, there is no doubt that the right hemisphere also contributes to speech perception (Jung-Beeman, 2005). However, the exact function of the right hemisphere in speech processing has yet to be elucidated. The auditory association cortex, which is situated in the posterior supratemporal plane, has mostly been associated with a variety of auditory functions (Warren *et al.,* 2005), amongst others with speech melody processing (Gandour *et al.,* 2004a; Homae *et al.,* 2006; Meyer *et al.,* 2002c; Meyer *et al.,* 2004b). Thus, one might hypothesize that speech rhythm processing, being classified as a prosodic aspect of speech might also be related to a right temporal lobe function, more specifically the posterior superior temporal gyrus (pSTG). Support for this reasoning is provided by a current model of functional lateralization in language processing – the "asymmetric sampling in time" (AST) hypothesis (Nicholls, 1996; Poeppel, 2003; van Wassenhove *et al.,* 2007). This model assumes that auditory fields in the two hemispheres prefer different temporal integration windows. Basically, the AST suggests that perception of rapidly changing cues (~40 Hz) preferentially drives the left hemisphere, the right auditory cortex is better adept at processing of slowly changing acoustic cues (~4 Hz) available in spoken language. The model further proposes that the posterior portion of auditory association cortex, in particular the planum temporale (PT), is the candidate region that accommodates this temporal processing device. Akin to speech melody, speech rhythm involves the integration of rhythmic speech elements

throughout a longer period of time such as the duration of a sentence. Thus we hypothesize that speech rhythm perception should be considered a suprasegmental process that is preferentially supported by brain structures of the right hemisphere.

It has been demonstrated by auditory functional imaging studies, that the pattern of hemodynamic responses in frontal (Hsieh *et al.*, 2001b) and in temporal cortical areas (Brechmann and Scheich, 2005c; Plante *et al.,* 2002) depends not only on the particular stimulus class but also on the task performed. Notably in the temporal cortex the hypothesis of task dependent functional lateralization is sustained by anatomical and functional plastic changes observed due to environmental and training influences (Josse *et al.,* 2003b; Luders *et al.,* 2004; Tervaniemi and Hugdahl, 2003a). Therefore, assuming that lateralization in the temporal cortex might be highly influenced by top-down modulation processes (Tervaniemi and Hugdahl, 2003b) we find it highly important to assess speech rhythm processing under different processing conditions.

The goal of the present study was to identify specific brain structures involved in the performance of a speech rhythm task. Furthermore, the study aimed to address the question of implicit (stimulus-driven) vs. explicit (task-driven) brain activation in speech rhythm perception.

Material and Methods

We investigated the processing of speech rhythm by using natural pseudo-sentences spoken with different rhythmic patterns. Haemodynamic responses were measured during explicit (task driven) and implicit (stimulus driven) auditory processing.

Participants

Twenty-three subjects (8 woman, mean age 27.4, SD= 9.3) participated in this study. One subject had to be excluded from the analysis due to a performance rate of below 63.8 percent correct answers. All participants

were right-handed according to the Annett-Handedness-Questionnaire (Annett, 1992). They were native speakers of Swiss German with no history of neurological, major medical, psychiatric or hearing disorders. Furthermore, subjects gave written consent in accordance with procedures approved by the local ethics committee and were paid for participation.

Stimuli

As mentioned in the introduction rhythmic distinctions can be found between languages. German is considered a stress-timed language that means it is the characteristic temporal distribution of accented syllables which constitutes its speech rhythm. This temporal distribution of accented syllables where accents are marked by modulation of pitch or intensity is not exactly isochronous in the spoken language, rather it is modulated by syllable lengthening and pauses. Notwithstanding, the German language shows a strong tendency towards isochrony (Noel Aziz Hanna, 2003; Völtz, 1991). This is especially evident in verse where speech becomes metric. We can conclude that spoken German can be described on a continuum between metric and non-metric, whereas verse is often metric. For the purpose of this study we decided to use sentences with an exaggerated German speech rhythm, i.e. “metrically” (isochronously) spoken sentences, as well as sentences with normal conversational speech rhythm, later referred to as “non-metrically” spoken sentences.

The stimulus material comprised a total of 144 German pseudo-sentences with natural speech rhythm manipulations. Pseudo-sentences follow the correct German syntax and phonology containing phontactically legal pseudo-words instead of content words. The sentences followed either a “non-metric” conversational speech rhythm or a “metrical”, isochronous, speech rhythm. The latter is supposed to represent an exaggerated form of German speech rhythm.

The metrically spoken sentences followed a regular meter (i.e. jambs, trochees, dactyls), with stressed syllables following each other isochronously. Non-metrically spoken sentences followed an irregular meter (i.e. jambs or

trochees with a dactyl interposed between two metrical feet), with a normal conversational flow of speech (Tab1).

Examples:

"Der Speiter pongt den spiten Galtung"	= „metric"
x ¦ X x ¦ X x ¦ X x ¦ X x ¦	
„ Der Jüfele knelt den furten Pflaster"	= „non-metric"
x ¦ X x x ¦ X x ¦ X x ¦ X x ¦	

Additionally, pseudosentences were recorded with the intonation contour of either a question or a statement balanced over the two rhythmic conditions. An active auditory control condition consisting of isochronous syllables (e.g. "da de di do du"; n=36) was constructed. Furthermore, a total of 40 nullevents as baseline condition were randomly included in the time course of the experiment which is a standard procedure in event-related fMRI.

The pseudosentences were constructed according to the phonotactical rules of the German language and controlled for syntactic differences across experimental conditions (Friederici *et al.,* 2000). All stimulus items were recorded from a trained German speaker and were normalized in amplitude to 70 percent of the loudest signal in a stimulus item. Then, all stimulus items were analyzed by means of the PRAAT speech editor (Boersma and Weenink, 2000) and controlled for duration and intensity on a root-means-square based measure.

Experimental Groups / Task

Subjects were randomly assigned to two different experimental groups each group having to perform a different categorization task on the same set of stimuli (Tab1). The "explicit processing" group (n=12) had to judge, whether they had heard a metric or non-metric pseudo-sentence (rhythm task). The "implicit processing" group (n=12) had to judge whether they had heard a statement or question (prosody task). Subjects indicated their

response by pressing a button with the index and middle finger respectively of their right hand. Subjects of the implicit processing group were naïve with respect to the rhythmic manipulation and did not detect a rhythmic difference between the stimuli throughout the experiment as confirmed by a short debriefing after the experiment. No feedback was given during the experiment.

Table 1: Experimental conditions

Processing modus / sentence type	explicit (n=12)	implicit (n=12)
Isochronous (n = 72)	**question** / **statement**	**question** / **statement**
non-isochronous (n = 72)	**question** / **statement**	**question** / **statement**

Procedure

In a short training session conducted prior to the fMRI experiment, subjects were made familiar with the task. We used an MR-compatible piezoelectric auditory stimulation system incorporated into standard Philips headphones for binaural stimulus delivery. Stimuli were presented in pseudorandom order binaurally. A total of 220 trials (144 pseudosentences, 36 active auditory control items, 40 nullevents) were presented in two sessions with a break in between the sessions. Before each stimulus presentation a fixation cross was presented for 500 ms.

fMRI Design

We implemented a clustered sparse temporal acquisition technique that combines the principles of a sparse temporal acquisition with a clustered acquisition of three consecutive volume scans per trial (Schmidt C.F. *et al.,* 2007). We collected three consecutive volumes in order to cover the peak of the event-related hemodynamic signal.

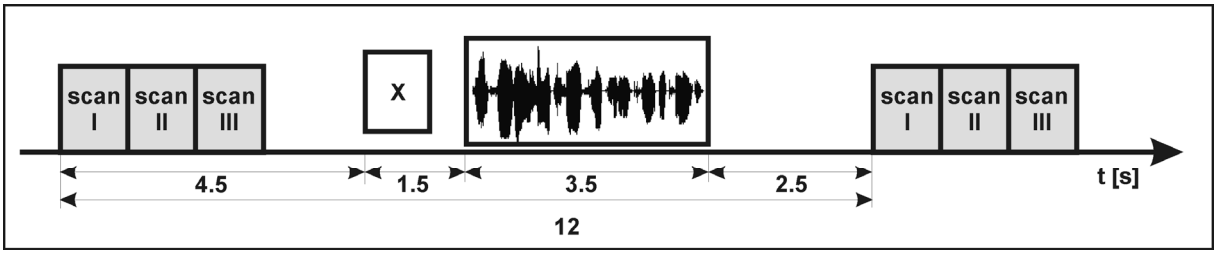

Figure 1: Acquisition scheme

Data acquisition

Data were collected using a Philips Intera 3T whole body MR unit (Philips Medical System Best, The Netherlands) equipped with an eight-channeled Philips SENSE head coil. Functional time series were collected from 16 transverse slices covering the entire peri-sylvian cortex with a spatial resolution of 2.7 x 2.7 x 4 mm using a Sensitivity Encoded (SENSE) (Pruessmann *et al.,* 1999) single-shot gradient-echo planar sequence (acquisition matrix 80 x 80 voxels, SENSE acceleration factor R = 2, FOV = 220 mm, TE = 35 ms and inter-slice gap 2 mm). Three volumes, each volume consisting of 16 axial slices, were acquired per trial with each a Tacq=1000 ms, θ = 68o (decay sampling) and 12 s inter cluster interval (ICI). Additionally, we obtained one whole-head EPI volume with 28 transversal slices (Tacq= 3000 ms, θ = 82o) but otherwise the identical scan parameters as for the functional time series to assist in the spatial normalization of the functional data. Furthermore, we collected a standard 3D T1-weighted scan for anatomical reference with 1 x 1 x 1.5 mm spatial resolution (180 axial slices, acquisition matrix 224 x 224 voxels, TE = 2.3 ms, TR = 20 ms, θ = 200).

Data analysis

Behavioural data

During the experiment the behavioral performance of the two experimental groups were measured. Due to technical problems, the performance data of 4 subjects could not be recorded. Data were corrected for outliers (>2 std above or below mean value). Behavioral measures were aggregated by participants and conditions. As a measure for accuracy of discrimination, the mean percentage of correct answers over all experimental

conditions (without baseline) was calculated. An independent sample t-test was performed to identify group differences.

fMRI data

To account for different T1 saturation effects in subsequent volumes, we subjected the three volume scans collected during each cluster to three separate analyses during pre-processing and individual statistics. The functional imaging data processing was carried out using MATLAB 6.5 (Mathworks Inc., Natiek, MA, USA) and the software package SPM99 (Wellcome Department of Cognitive Neurology, London, UK, http://www.fil.ion.ucl.ac.uk/spm/). Functional data were realigned to the first volume, corrected for motion artifacts and normalized into standard stereotactic space (voxel size 2 x 2 x 2 mm, template provided by the Montreal Neurological Insitute (Evans *et al.,* 1992)). For spatial smoothing we applied an isotropic Gaussian kernel of 8 mm full-width-at-half-maximum. Low-frequency drifts were removed using a temporal high-pass filter (mean cut-off of 307 seconds).

Statistical evaluation was based on a least-square estimation using the general linear model for serially autocorrelated observations, performed separately on each voxel (Friston *et al.,* 1995). Single trials were treated as epochs and modeled by means of a box car function. Conditions were compared by calculating contrasts between conditions of interest for each participant and time point of acquisition. We decided to report the second acquisition time as they are supposed to reflect the amplitude peak of the hemodynamic response as it has been done formerly (Bunzeck *et al.,* 2005; Meyer *et al.,* 2007). Contrast images were submitted to a second level group analysis. Random-effects analysis consisted of between and within group comparison. The between-group analysis consisted of a two-sampled t-test using “metric and non-metric sentences>auditory control” contrasts. The within-group analysis consisted of a paired-t-test using “metric sentences>auditory control” and “non-metric sentences>auditory control” contrasts. Only clusters of significant size ($p < 0.05$, corrected for multiple

comparisons) and with a T-threshold of 4.02 ($p < 0.001$, uncorrected for multiple comparisons) were reported (Worsley *et al.*, 1996).

Results

Behavioral results

The implicit processing group performed better and faster than the explicit processing group. 80.25% of the answers were correct for the explicit processing group and 98.6% for the implicit processing group. An independent sample t-test performed to identify group differences revealed significant difference in response accuracy ($t(18) = -5.28$; $p < 0.000$).

fMRI results

The present paper addressed two main issues. First we were interested in the neurofunctional correlate related to the performance of the speech rhythm task. Second we aimed to analyze the brain areas involved in the processing of exaggerated German speech rhythm in an explicit and an implicit processing condition.

Performance of the speech rhythm task

Between-group comparison between explicit and implicit processing of all sentences (Fig 2, Tab 2) revealed activation in the supplementary motor area (SMA) for the explicit processing group compared to the implicit processing group. This activation was found bilaterally, although the peak activation was in the right hemisphere. The cluster extended into the medial part of the superior frontal gyrus and the right cingulate gyrus. Furthermore, this contrast revealed bilateral activation in the left and right insula encroaching onto the neostriatum, namely the Putamen in the right hemisphere. In the right hemisphere we found an additional activation in the inferior frontal gyrus (pars opercularis).

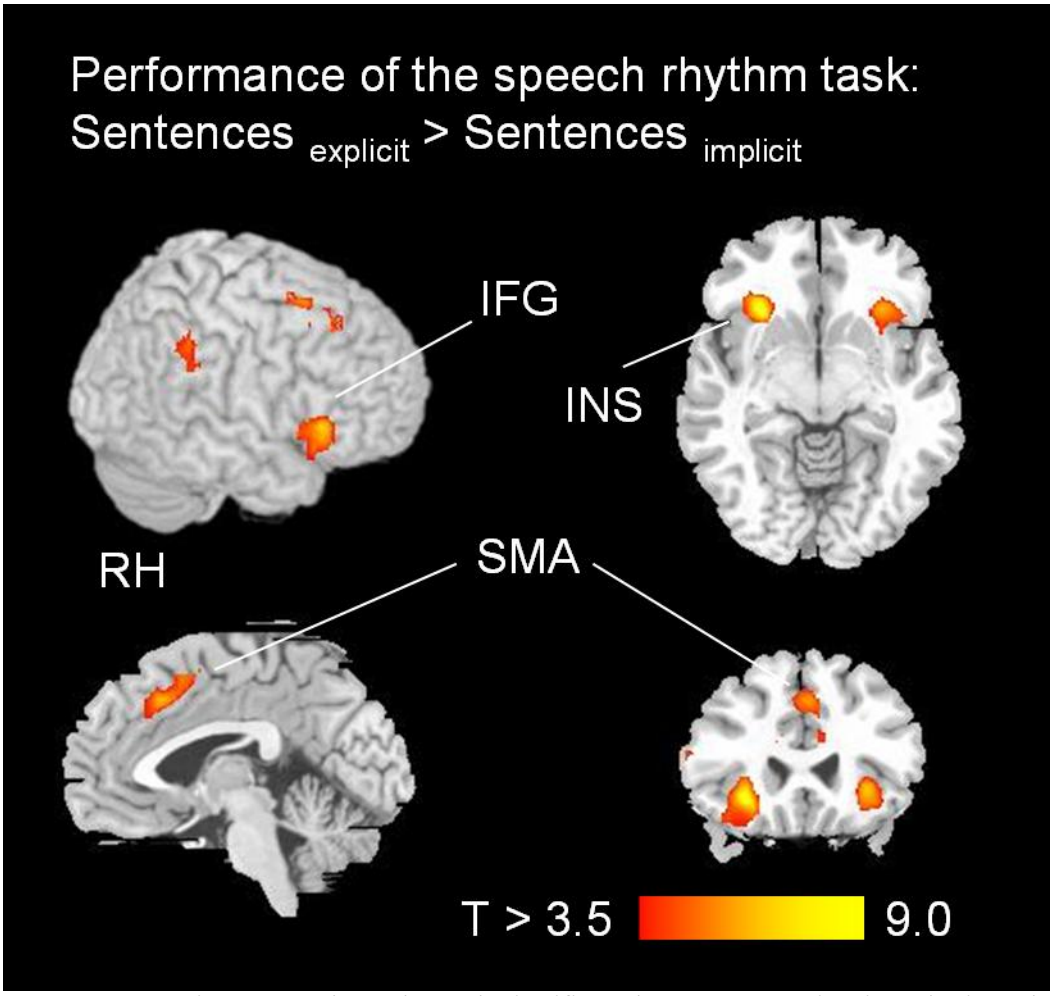

Figure 2: Brain areas that showed significantly greater activation during the performance of the rhythm task (explicit) as compared to the implicit processing condition. The figure shows resulting T-maps of the two-sampled t-test using sentences vs. auditory control contrast for the explicit as compared to the implicit processing condition (up left z= -3, down left x= -2, down right y = 25).The top-right image shows the activations projected on the cortical surface. Each cluster is thresholded at T = 4.02 ($p<0.001$, uncorrected) and cluster corrected for multiple comparisons at $p<0.05$ (Worsley et al. 1996).

Explicit processing of metrical speech rhythm

To reveal the effect of speech rhythm we compared metrically spoken sentences to non-metrically spoken sentences in the explicit processing condition (Fig 3, Tab 3). We observed statistically significant activation in right superior temporal gyrus extending into the supramarginal gyrus and the Parietal Operculum.

Implicit processing of metrical speech rhythm

The contrast between metrically spoken and non-metrically spoken sentences in the implicit processing condition (Fig 3, Tab 4) revealed activation in the left hemisphere, namely the supramarginal gyrus and the superior temporal gyrus extending into the Parietal Operculum for the metrically spoken sentences compared to non-metrically spoken sentences.

Taken together, direct contrast between metrical and non-metrical sentences revealed a substantial effect observed for the processing of metrically spoken speech, which was differentially lateralized in the explicit as compared to the implicit processing condition. Furthermore, the performance of the speech rhythm task relied on activation of the supplementary motor area (SMA) as well as the insula (INS) bilaterally and the right inferior frontal gyrus (IFG).

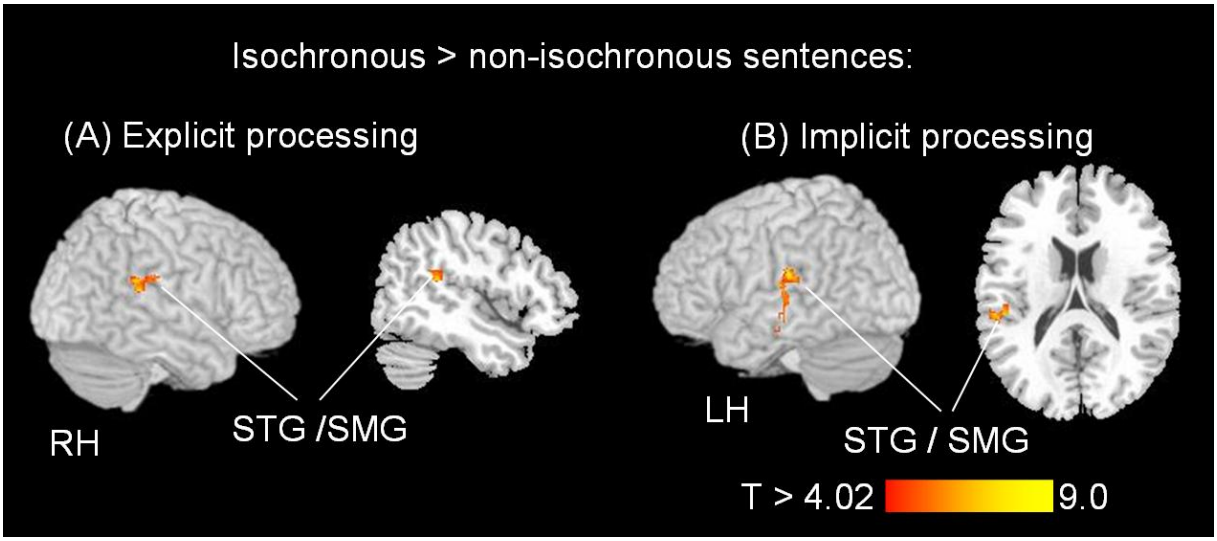

Figure 3: Brain areas showing significantly greater activation during the processing of metrical compared to non-metrical sentences in (A) the explicit processing condition (right x = 48) and (B) the implicit processing condition (right *z* = 20). For both the explicit and implicit processing condition the activation in the sagital view is projected on the lateral convexity of cortical surface. Each cluster is thresholded at T = 4.02 ($p<0.001$, uncorrected) and cluster corrected for multiple comparisons at $p<0.05$ (Worsley et al. 1996).

Table 2: Performance of the speech rhythm task

two-sampeld t-test:

(metric and non-metric sentences>auditory control) expl > (metric and non-metric sentences>auditory control) imp

	Left hemisphere				Right hemisphere			
Anatomical description	p value*	T value	cluster size	x,y,z [mm]	p value*	T value	cluster size	x,y,z [mm]
INS extending into NS					0.000	7.57	212	27 21 -9
IFG (pars opercularis)					0.000	6.89	168	51 12 12
SMG					0.000	6.38	257	42 -42 36
INS extending into IFG (pars orbitalis/triangularis)	0.000	6.3	176	-30 24 -3				
SMA and cingulate gyrus extending into SFG	0.000	6.12	249	0 21 42				
IPL	0.041	4.89	46	-39 -42 42				

* p-values corrected for entire volume on cluster-level (k=**45**)

Table 3: Explicit metrical speech rhythm processing

paired t-test: (metric sentences>auditory control) > (non-metric sentences>auditory control)

	Left hemisphere				Right hemisphere			
Anatomical description	p value*	T value	cluster size	x,y,z [mm]	p value*	T value	cluster size	x,y,z [mm]
STG extending into SMG / PaOp					0.002	6.72	49	48 -39 18

* p-values corrected for entire volume on cluster-level (k = **27**)

Table 4: Implicit metrical speech rhythm processing

paired t-test: (metric sentences>auditory control) > (non-metric sentences>auditory control)

	Left hemisphere				Right hemisphere			
Anatomical description	p value*	T value	cluster size	x,y,z [mm]	p value*	T value	cluster size	x,y,z[mm]
SMG extending into STG / PaOp	0.000	7.45	75	-57 -30 21				

* p-values corrected for entire volume on cluster-level (k=**27**)

Legend:

INS = Insula, NS = neostriatum (caudate/head, putamen), IPL = Inferior parietal lobe, Ppar = Planum parietale, RO = Rolandic Operculum, STG = Superior temporal gyrus, SFG = Superior frontal gyrus, OG = occipital gyrus, PP = planum polare, PT = planum temporale, SMA = supplementary motor area, IPG = Inferior parietal gyrus, SMG = Supplementary motor area, PaOp = parietal operculum

Discussion

The main goal of the study was to identify brain areas related to speech rhythm processing, that is, the performance of a speech rhythm task. A further objective was to determine possible differences between explicit and implicit rhythm perception. In the following section we will first discuss the brain activation evoked by the performance of the speech rhythm task. Then we will discuss the differential effects found for metric processing as compared to non-metric processing in both the explicit and subsequently in the implicit processing condition.

Performance of the speech rhythms task

A clear effect related to the explicit performance of the rhythm task was observed in the mesial premotor cortex (SMA-pre/proper), in contrast to the implicit processing group. We suggest that this finding gives rise to the notion of a general function of the mesial premotor cortex which is related to timing processes in auditory perception as described in the following section. Considering other recent findings this result speaks for a multimodal function in temporal perception of the SMA.

The SMA has frequently been reported to be involved in motor processes such as finger movement or speech production (Riecker *et al.,* 2005). In this research field, the SMA is convergently divided into pre-SMA, associated with motor preparation, and SMA-proper, associated with movement execution, encoding and retrieval of motor sequences (Cunnington *et al.,* 2005; Kansaku *et al.,* 2005; Koeneke *et al.,* 2004; Lau, 2006; Lewis *et al.,* 2004; Lutz *et al.,* 2000a). However, there is evidence that this involvement of the SMA-proper as well as the pre-SMA in motor processes might be due to a specific function in the timing of motor actions, a function which might not only be restricted to motor processing. Both areas have been found to show an increasing relation between complexity, that is temporal complexity, of motor sequences and activation (Bengtsson *et al.,* 2005; Lewis *et al.,* 2004; Lutz *et al.,* 2000b; Riecker *et al.,* 2006).

The assumption of a temporal processing function of the SMA is sustained by evidence from perception research in the visual domain, which described a correlation between attention to time and activity in the SMA (Coull, 2004; Coull *et al.,* 2004; Lewis and Miall, 2003a; Nobre and O'Reilly, 2004). A similar finding has been reported in the sensorimotor domain showing an involvement of the SMA in time estimation after a specific stimulation (Macar *et al.,* 2004). Interpreting the above mentioned studies in the direction of a multimodal involvement of the SMA in time perception (Pastor *et al.,* 2006; Schubotz *et al.,* 2000), our study poses an enhancement of this assumption in that we used acoustic stimuli and found a very specific activation in the SMA related to temporal perception. Furthermore, our results found in the contrast between sentences and baseline condition, clearly speak for a task-driven effect of sequential timing perception.

An alternative interpretation of the SMA activation could be given in terms of strategy mechanisms. It is possible that the SMA activation relates to spontaneous and intuitive synchronization of rhythm perception and body movements. That is, subjects intuitively "feel" the rhythm in their body while listening to the specific sentences. This assumption is sustained by reports showing that auditory cues are very effective in facilitating accurate rhythmic body movements (Patel *et al.,* 2005) as well as by comments made by subjects after the performance of the experiment. This sensation could be mediated by motor cortices such as the SMA. However, spontaneous motor synchronization mechanisms must be a topic of further investigations.

Furthermore, our results show an extensive activation in the insula of both hemispheres and in the right inferior frontal gyrus, specifically the right pars opercularis. Besides its function in motor (Jantzen *et al.,* 2005) and speech production processes (Ackermann and Riecker, 2004; Blank *et al.,* 2002; Bohland and Guenther, 2006), specifically in motor timing, the insula has repeatedly been found in various auditory processes (Bamiou *et al.,* 2003; Giraud *et al.,* 2004). Of note are auditory temporal processing (Ackermann *et al.,* 2001a; Lewis *et al.,* 2000; Lewis and Miall, 2003b), speech melody perception (Meyer et al. 2002), and auditory feedback and pitch

control (Zarate and Zatorre, 2005) which have recently been reported. Moreover, functional lateralization in the insulae has been suggested on the basis of temporal acoustic characteristics (Ackermann *et al.,* 2001b) as well as on the function of pitch-related stimuli (Wong *et al.,* 2004). Very recently Bamiou et al. (Bamiou *et al.,* 2006) gave strong support to the notion of a crucial role of the insula in temporal processing, in particular, temporal resolution and sequencing. Thus, there is an evolving view that the insula is crucial in auditory timing perception. Our processing task involves the perception of the specific accent patterns, that is, the sequencing of syllables. Therefore, our finding of an insular involvement in the sequencing task is not surprising and in line with the above mentioned hypothesis about insular functions.

Of particular note is the activation in the right insula extending into the basal ganglia (Putamen). Lesion studies as well as imaging studies revealed an involvement of the basal ganglia in the processing of suprasegmental speech cues, namely emotional speech melody in speech production (Pell *et al.,* 2006) and emotional speech melody perception (Kotz *et al.,* 2003; Meyer *et al.,* 2004a). This data expands the function of the basal ganglia over the commonly reported functions in motor processing. Furthermore, lesion studies have found that patients with a basal ganglia insult also show symptoms of impaired prosodic functions (Van Lancker Sidtis D. *et al.,* 2006). Thus, our data suggest that the involvement of basal ganglia in the processing of suprasegmental speech cues might not be limited to melodic aspects of speech but be necessary for the perception of rhythmic aspects of speech too.

We want to put an additional emphasis on the activity observed in the right inferior frontal gyrus. The right IFG has repeatedly been reported in addition to left frontal brain areas in speech processing (Constable *et al.,* 2004b; Michael *et al.,* 2001b; Ni *et al.,* 2000b) and associated with increasing sentence complexity (Michael *et al.,* 2001a), or syntax processing (Meyer *et al.,* 2000b; Moro *et al.,* 2001; Ni *et al.,* 2000a; Poldrack *et al.,* 2001). This activation was often found to be related to the performance of specific speech tasks (Chou *et al.,* 2006; Meyer *et al.,* 2000a) or even associated with task

related attention mechanisms (Hashimoto *et al.,* 2000; Ruff *et al.,* 2003). However, there is also a considerable amount of research which has associated the right IFG with specific acoustic perception, i.e. pitch processing. Pitch processing plays a crucial role in speech processing of melodic languages (Gandour *et al.,* 2003). In non-melodic languages pitch becomes important in the context of speech melody perception (Constable *et al.,* 2004a; Meyer *et al.,* 2002b).

The perception of accent patterns in the German language is indivisible from pitch and intensity patterns, since an accent is related to the relative change of both acoustic characteristics. Therefore, we strongly believe that in our study the activation in the right IFG must be interpreted in relation to the integration of accent patterns as they are crucial in the categorization of different speech rhythms. Activation in the right IFG has been found during pitch processing in both non-linguistic (Hsieh *et al.,* 2001a; Humphries *et al.,* 2001; Liu *et al.,* 2006) and linguistic contexts (Gandour *et al.,* 2004b; Meyer *et al.,* 2002a; Wildgruber *et al.,* 2006). As Chinese inept subjects process Chinese stimuli with by focusing on the pitch pattern independently of its linguistic relevance, our subjects might have processed the stimuli with focus on the specific accent pattern, since they had a speech rhythm task to do. We believe that our data provide strong evidence, that the right IFG can be selectively activated in language processes involving higher suprasegmental speech processing, more specifically the detection of pitch changes.

Explicit and implicit processing of metrical speech rhythm

We analyzed the effect of speech rhythm, metric vs. non-metric sentences, in both the explicit and the implicit processing condition. This analysis is based on recent findings of differentially lateralized effects in auditory processing depending on the context of the stimulus perception (Brechmann and Scheich, 2005b) as well as on findings reporting a top-down modulation on auditory processing (Tervaniemi and Hugdahl, 2003c).

In the explicit processing condition we measure task-driven processing of metrical speech. Interestingly, this process leads to temporal activation in

the right hemisphere, namely activation in the posterior part of the right STG (specifically planum temporale, PT) extending into the supramarginal gyrus, for the metric as compared to the non-metric sentences. Most theories of the functional role of the PT now assume that the PT in the left and the right hemisphere is involved in a parameter specific segregation of auditory sounds leading to differentially lateralized processes of the two hemispheres. Two different parameter-specific concepts of the PT are found in the literature, however they are widely considered to be two sides of the same coin. One concepts posits that the posterior part of the STG is sensitive to spectrotemporal auditory processing and associates the right pSTG with spectral and the left pSTG with temporal auditory processing (Griffiths and Warren, 2002; Menon *et al.,* 2002; Meyer *et al.,* 2005; Zatorre and Belin, 2001). A complimentary model, the AST, specifies the potential functions of the pSTG slightly differently (Poeppel, 2003). This model proposes that the left pSTG is preferentially driven by rapidly changing acoustic cues, i.e. formant transitions, while right pSTG is more amenable to slowly changing acoustic cues, i.e. intonation contours. The perception of suprasegmentally modulated speech rhythm that enfolds in the time range of syllables (150-200 ms) should be therefore more likely supported by the right posterior temporal plane. Thus we state that our findings can be explained by the AST of functional lateralization. However, our results have a second highly important implication. We find the right hemisphere to be involved in the processing of metrical speech rhythm in the explicit processing condition only. The same contrast in the implicit processing condition does not show right hemispheric activation. Therefore, our results additionally provide strong evidence for the influence of top-down modulation on lateralization effects in the auditory cortex (Brechmann and Scheich, 2005a; Tervaniemi and Hugdahl, 2003d).

In the implicit processing condition we measure pure stimulus-driven processing of speech rhythm. Furthermore, as confirmed after the completion of the experiment subjects were not aware of the rhythmic manipulation within the sentences. Surprisingly, the contrast between

metric and non-metric sentences revealed significant activation in the left supramarginal gyrus extending into the left posterior superior temporal gyrus (STG), more specifically the planum temporale (PT) and extending in the left Parietal Operculum for the metric condition as compared to the non-metric condition. The extent of activation observed in the left posterior STG encroached onto the upper-posterior part of the Sylvain fissure and the supramarginal gyrus. The extension of activity over these cortical areas is not surprising. It has been shown that subjects with a larger left PT tend to show a larger leftward functional asymmetry of several perisylvian areas, namely the inferior parietal lobule outside the supra-marginal gyrus, Heschl's gyrus, the rolandic operculum, and the temporal pole (Josse *et al.,* 2003a). This finding suggests possible functional common ground between these cortex areas. Hickok and collegues (Hickok *et al.,* 2003) first described activation in the parietal operculum and in the adjacent posterior part of the Sylvian fissure, at the parietal-temporal boundary for both speech listening tasks and covert rehearsal of spoken language. The authors suggested that this area might be involved in auditory-motor activity similar to the visual-motor response properties in the dorsal visual stream (Rizzolatti *et al.,* 1997). In the context of our study it is important to notice that musical meter, as the whole-numbered (i.e. isochronous) sequence of accented acoustic events, is the characteristic feature of music which relates to body movements (Lerdahl and Jackendoff, 1983). Very recently Chen et al. (Chen *et al.,* 2006) investigated the neural correlate of a synchronization process of body movements using the meter of an auditory cue. The authors found the left planum temporale, the right posterior STG and bilateral dorsal premotor cortex to covariate with metric salience and thus to facilitate the synchronization of motor actions to auditory cues. Therefore, we suggest that listening to exaggerated speech rhythm in the implicit processing condition leads to involuntary auditory-motor interactions due to the salience of metrically spoken sentences.

Conclusion

We found the SMA, right IFG, bilateral insulae as well as the right basal ganglia to be involved in sequential timing perception related to the performance of a rhythmic categorization task on speech stimuli.

Our data provide evidence for an involvement of the supplementary motor area in the perception of time intervals across modalities and for a specific task-related function of the right inferior frontal gyrus in the processing of accent patterns in speech.

Furthermore, we observed a clear effect of speech rhythm which was lateralized to the right auditory cortex in an explicit, task-driven processing condition. This observation provides evidence in support of our hypothesis which predicts that suprasegmental speech rhythm processing occurs over a longer temporal integration window. However, this activation does not occur in implicit processing conditions. This finding indicates that right secondary auditory cortex activity can be modulated by top-down processing and can be selectively recruited depending on task demands. The same contrast in an implicit, stimulus-driven processing condition evoked differentially lateralized brain activation in the left hemisphere. This finding is assumed to reflect an involuntary auditory-motor interaction effect.

Acknowledgments

This work was supported by the Swiss National Foundation (SNF-grant 3200B0-105877 conferred to M. Meyer). We are grateful to the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig for assistance in stimulus recording. We specially thank Dr. Dorothea Weniger from the University Hospital of Zurich for helpful comments on previous versions of this manuscript. Furthermore, we thank Dr. Conny F. Schmidt from the Institute for Biomedical Technology, ETH and University of Zurich, Switzerland, for assistance in the data acquisition.

Reference List

Ackermann, H., Riecker, A., 2004. The contribution of the insula to motor aspects of speech production: a review and a hypothesis. Brain Lang 89, 320-328.

Ackermann, H., Riecker, A., Mathiak, K., Erb, M., Grodd, W., Wildgruber, D., 2001a. Rate-dependent activation of a prefrontal-insular-cerebellar network during passive listening to trains of click stimuli: an fMRI study. Neuroreport 12, 4087-4092.

Ackermann, H., Riecker, A., Mathiak, K., Erb, M., Grodd, W., Wildgruber, D., 2001b. Rate-dependent activation of a prefrontal-insular-cerebellar network during passive listening to trains of click stimuli: an fMRI study. Neuroreport 12, 4087-4092.

Annett, M., 1992. 5 Tests of Hand Skill. Cortex 28, 583-600.

Bamiou, D.E., Musiek, F.E., Luxon, L.M., 2003. The insula (Island of Reil) and its role in auditory processing. Literature review. Brain Res.Brain Res.Rev. 42, 143-154.

Bamiou, D.E., Musiek, F.E., Stow, I., Stevens, J., Cipolotti, L., Brown, M.M., Luxon, L.M., 2006. Auditory temporal processing deficits in patients with insular stroke. Neurology 67, 614-619.

Bengtsson, S.L., Ehrsson, H.H., Forssberg, H., Ullen, F., 2005. Effector-independent voluntary timing: behavioural and neuroimaging evidence. Eur.J.Neurosci. 22, 3255-3265.

Blank, S.C., Scott, S.K., Murphy, K., Warburton, E., Wise, R.J., 2002. Speech production: Wernicke, Broca and beyond. Brain 125, 1829-1838.

Boersma, P., Weenink, D., 2000. PRAAT: Doing phonetics by computer. Instiute of phonetic Sciences, University of Amsterdam.

Bohland, J.W., Guenther, F.H., 2006. An fMRI investigation of syllable sequence production. Neuroimage.

Bolinger, D., 1965. Pitch accent and sentence rhythm, Forms of English: accent, morpheme, order. Harvard University Press, Cambridge MA.

Brechmann, A., Scheich, H., 2005a. Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cereb.Cortex 15, 578-587.

Brechmann, A., Scheich, H., 2005b. Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cereb.Cortex 15, 578-587.

Brechmann, A., Scheich, H., 2005c. Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cereb.Cortex 15, 578-587.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H.J., Jancke, L., 2005. Scanning silence: mental imagery of complex sounds. Neuroimage. 26, 1119-1127.

Chen, J.L., Zatorre, R.J., Penhune, V.B., 2006. Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. Neuroimage.

Chou, T.L., Booth, J.R., Burman, D.D., Bitan, T., Bigio, J.D., Lu, D., Cone, N.E., 2006. Developmental changes in the neural correlates of semantic processing. Neuroimage. 29, 1141-1149.

Constable, R.T., Pugh, K.R., Berroya, E., Mencl, W.E., Westerveld, M., Ni, W., Shankweiler, D., 2004b. Sentence complexity and input modality effects in sentence comprehension: an fMRI study. Neuroimage. 22, 11-21.

Constable, R.T., Pugh, K.R., Berroya, E., Mencl, W.E., Westerveld, M., Ni, W., Shankweiler, D., 2004a. Sentence complexity and input modality effects in sentence comprehension: an fMRI study. Neuroimage. 22, 11-21.

Coull, J.T., 2004. fMRI studies of temporal attention: allocating attention within, or towards, time. Brain Res.Cogn Brain Res. 21, 216-226.

Coull, J.T., Vidal, F., Nazarian, B., Macar, F., 2004. Functional anatomy of the attentional modulation of time estimation. Science 303, 1506-1508.

Cunnington, R., Windischberger, C., Robinson, S., Moser, E., 2005. The selection of intended actions and the observation of others' actions: A time-resolved fMRI study. Neuroimage.

Curtin, S.M.T.H.C.M.H., 2006. Stress changes the representational landscape: Evidence from word segmentation. Cognition.

Cutler, E.A., Norris, D., 1988. The role of strong syllables in segmentation for lexical access. In: Altmann G.T.M. (Ed.), Psycholinguistics: Critical Concepts, pp. 157-177.

Evans, A.C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W., Milot, S., Meyer, E., Bub, D., 1992. Anatomical mapping of functional activation in stereotactic coordinate space. Neuroimage. 1, 43-53.

Frazier, L., Carlson, K., Clifton, C., Jr., 2006. Prosodic phrasing is central to language comprehension. Trends Cogn Sci. 10, 244-249.

Friederici, A.D., Meyer, M., von Cramon, D.Y., 2000. Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. Brain Lang 75, 289-300.

Friedrich, C.K., Kotz, S.A., Friederici, A.D., Alter, K., 2004a. Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. Brain Res.Cogn Brain Res. 20, 300-308.

Friedrich, C.K., Kotz, S.A., Friederici, A.D., Alter, K., 2004b. Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. Brain Res.Cogn Brain Res. 20, 300-308.

Friston, K.J., Holmes, A.P., Worsley, K.P., Poline, J.B., Frith, C.D., Frackowiak, R.S., 1995. Statistical parameter maps in functional imaging: A general linear approach. Hum.Brain Mapp. 2, 189-210.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., Lowe, M., 2004b. Hemispheric roles in the perception of speech prosody. Neuroimage. 23, 344-357.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., Lowe, M., 2004a. Hemispheric roles in the perception of speech prosody. Neuroimage. 23, 344-357.

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X., Tong, Y., 2003. Neural correlates of segmental and tonal information in speech perception. Hum.Brain Mapp. 20, 185-200.

Giraud, A.L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M.O., Preibisch, C., Kleinschmidt, A., 2004. Contributions of Sensory Input, Auditory Search and Verbal Comprehension to Cortical Activity during Speech Processing. Cereb.Cortex 14, 247-255.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348-353.

Hashimoto, R., Homae, F., Nakajima, K., Miyashita, Y., Sakai, K.L., 2000. Functional differentiation in the human auditory and language areas revealed by a dichotic listening task. Neuroimage 12, 147-158.

Hickok, G., Buchsbaum, B., Humphries, C., Muftuler, T., 2003. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J.Cogn Neurosci. 15, 673-682.

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., Taga, G., 2006. The right hemisphere of sleeping infant perceives sentential prosody. Neurosci.Res. 54, 276-280.

Hsieh, L., Gandour, J., Wong, D., Hutchins, G.D., 2001a. Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. Brain Lang 76, 227-252.

Hsieh, L., Gandour, J., Wong, D., Hutchins, G.D., 2001b. Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. Brain Lang 76, 227-252.

Humphries, C., Willard, K., Buchsbaum, B., Hickok, G., 2001. Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. Neuroreport 12, 1749-1752.

Jantzen, K.J., Steinberg, F.L., Kelso, J.A., 2005. Functional MRI reveals the existence of modality and coordination-dependent timing networks. Neuroimage. 25, 1031-1042.

Josse, G., Mazoyer, B., Crivello, F., Tzourio-Mazoyer, N., 2003b. Left planum temporale: an anatomical marker of left hemispheric specialization for language comprehension. Brain Res.Cogn Brain Res. 18, 1-14.

Josse, G., Mazoyer, B., Crivello, F., Tzourio-Mazoyer, N., 2003a. Left planum temporale: an anatomical marker of left hemispheric specialization for language comprehension. Brain Res.Cogn Brain Res. 18, 1-14.

Jung-Beeman, M., 2005. Bilateral brain processes for comprehending natural language. Trends Cogn Sci. 9, 512-518.

Kansaku, K., Muraki, S., Umeyama, S., Nishimori, Y., Kochiyama, T., Yamane, S., Kitazawa, S., 2005. Cortical activity in multiple motor areas during sequential finger movements: an application of independent component analysis. Neuroimage 28, 669-681.

Koeneke, S., Lutz, K., Wustenberg, T., Jancke, L., 2004. Long-term training affects cerebellar processing in skilled keyboard players. Neuroreport 15, 1279-1282.

Kotz, S.A., Meyer, M., Alter, K., Besson, M., von Cramon, D.Y., Friederici, A.D., 2003. On the lateralization of emotional prosody: an event-related functional MR investigation. Brain Lang 86, 366-376.

Lau, H., Rogers,.R.D., Passingham,.R.E., 2006. Dissociating response selection and conflict in the medial frontal surface. Neuroimage 29, 446-451.

Lee, C.S., Todd, N.P., 2004. Towards an auditory account of speech rhythm: application of a model of the auditory 'primal sketch' to two multi-language corpora. Cognition 93, 225-254.

Lehiste, I., 1977. Isochrony Reconsidered. Journal of Phonetics 5, 253-263.

Lerdahl, F., Jackendoff, R., 1983. An Overview of Hierarchical Structure in Music. Music Perception 1, 229-252.

Lewis, J.W., Beauchamp, M.S., DeYoe, E.A., 2000. A comparison of visual and auditory motion processing in human cerebral cortex. Cereb.Cortex 10, 873-888.

Lewis, P.A., Miall, R.C., 2003b. Brain activation patterns during measurement of sub- and supra-second intervals. Neuropsychologia 41, 1583-1592.

Lewis, P.A., Miall, R.C., 2003a. Brain activation patterns during measurement of sub- and supra-second intervals. Neuropsychologia 41, 1583-1592.

Lewis, P.A., Wing, A.M., Pope, P.A., Praamstra, P., Miall, R.C., 2004. Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. Neuropsychologia 42, 1301-1312.

Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K., Chen, C., 2006. Dissociation in the neural basis underlying Chinese tone and vowel production. Neuroimage. 29, 515-523.

Luders, E., Gaser, C., Jancke, L., Schlaug, G., 2004. A voxel-based approach to gray matter asymmetries. Neuroimage. 22, 656-664.

Lutz, K., Specht, K., Shah, N.J., Jancke, L., 2000a. Tapping movements according to regular and irregular visual timing signals investigated with fMRI. Neuroreport 11, 1301-1306.

Lutz, K., Specht, K., Shah, N.J., Jancke, L., 2000b. Tapping movements according to regular and irregular visual timing signals investigated with fMRI. Neuroreport 11, 1301-1306.

Macar, F., Anton, J.L., Bonnet, M., Vidal, F., 2004. Timing functions of the supplementary motor area: an event-related fMRI study. Brain Res.Cogn Brain Res. 21, 206-215.

McQueen, J.M., Otake, T., Cutler, A., 2001. Rhythmic cues and possible-word constraints in Japanese speech segmentation. Journal of Memory and Language 45, 103-132.

Menon, V., Levitin, D.J., Smith, B.K., Lembke, A., Krasnow, B.D., Glazer, D., Glover, G.H., McAdams, S., 2002. Neural correlates of timbre change in harmonic sounds. Neuroimage. 17, 1742-1754.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., 2002a. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum.Brain Mapp. 17, 73-88.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., 2002b. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum.Brain Mapp. 17, 73-88.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., von Cramon, D.Y., 2002c. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum.Brain Mapp. 17, 73-88.

Meyer, M., Baumann, S., Marchina, S., Jaencke, L., 2007. Haemodynamic responses in human mutlisensory and auditory association cortex to purely visual stimulation. BMC Neuroscience (in press).

Meyer, M., Friederici, A.D., von Cramon, D.Y., 2000a. Neurocognition of auditory sentence comprehension: event related fMRI reveals sensitivity to syntactic violations and task demands. Brain Res.Cogn Brain Res. 9, 19-33.

Meyer, M., Friederici, A.D., von Cramon, D.Y., 2000b. Neurocognition of auditory sentence comprehension: event related fMRI reveals sensitivity to syntactic violations and task demands. Brain Res.Cogn Brain Res. 9, 19-33.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D., von Cramon, D.Y., 2004a. Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain Lang 89, 277-289.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D., von Cramon, D.Y., 2004b. Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain Lang 89, 277-289.

Meyer, M., Zaehle, T., Gountouna, V.E., Barron, A., Jancke, L., Turk, A., 2005. Spectro-temporal processing during speech perception involves left posterior auditory cortex. Neuroreport 16, 1985-1989.

Michael, E.B., Keller, T.A., Carpenter, P.A., Just, M.A., 2001b. fMRI investigation of sentence comprehension by eye and by ear: modality fingerprints on cognitive processes. Hum.Brain Mapp. 13, 239-252.

Michael, E.B., Keller, T.A., Carpenter, P.A., Just, M.A., 2001a. fMRI investigation of sentence comprehension by eye and by ear: modality fingerprints on cognitive processes. Hum.Brain Mapp. 13, 239-252.

Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S.F., Fazio, F., 2001. Syntax and the brain: disentangling grammar by selective anomalies. Neuroimage. 13, 110-118.

Nazzi, T., Ramus, F., 2003a. Perception and acquisition of linguistic rhythm by infants. Speech Communication 41, 233-243.

Nazzi, T., Ramus, F., 2003b. Perception and acquisition of linguistic rhythm by infants. Speech Communication 41, 233-243.

Ni, W., Constable, R.T., Mencl, W.E., Pugh, K.R., Fulbright, R.K., Shaywitz, S.E., Shaywitz, B.A., Gore, J.C., Shankweiler, D., 2000b. An event-related neuroimaging study distinguishing form and content in sentence processing. J.Cogn Neurosci. 12, 120-133.

Ni, W., Constable, R.T., Mencl, W.E., Pugh, K.R., Fulbright, R.K., Shaywitz, S.E., Shaywitz, B.A., Gore, J.C., Shankweiler, D., 2000a. An event-related neuroimaging study distinguishing form and content in sentence processing. J.Cogn Neurosci. 12, 120-133.

Nicholls, M.E., 1996. Temporal processing asymmetries between the cerebral hemispheres: evidence and implications. Laterality. 1, 97-137.

Nobre, A.C., O'Reilly, J., 2004. Time is of the essence. Trends in Cognitive Sciences 8, 387-389.

Noel Aziz Hanna, P., 2003. Sprachrhythmus in Metrik und Alltagssprache. Wilhelm Fink Verlag, München.

Pastor, M.A., Macaluso, E., Day, B.L., Frackowiak, R.S., 2006. The neural basis of temporal auditory discrimination. Neuroimage 30, 512-520.

Patel, A.D., Iversen, J.R., Chen, Y., Repp, B.H., 2005. The influence of metricality and modality on synchronization with a beat. Exp.Brain Res. 163, 226-238.

Pell, M.D., Cheang, H.S., Leonard, C.L., 2006. The impact of Parkinson's disease on vocal-prosodic communication from the perspective of listeners. Brain Lang 97, 123-134.

Pike, K.L., 1945. The intonation of American English. University of Michigan Press, Ann Arbor.

Plante, E., Creusere, M., Sabin, C., 2002. Dissociating sentential prosody from sentence processing: activation interacts with task demands. Neuroimage. 17, 401-410.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication 41, 245-255.

Poldrack, R.A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., Gabrieli, J.D., 2001. Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. J.Cogn Neurosci. 13, 687-697.

Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., 1999. SENSE: Sensitivity encoding for fast MRI. Magnetic Resonance in Medicine 42, 952-962.

Ramus, F., Nespor, M., Mehler, J., 1999. Correlates of linguistic rhythm in the speech signal. Cognition 73, 265-292.

Riecker, A., Kassubek, J., Gröschel, K., Grodd, W., Ackermann, H., 2006. The cerebral control of speech tempo: Opposite relationship between speaking rate and BOLD signal chagnes at striatal and cerebellar structures. Neuroimage 29, 46-53.

Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., Ackermann, H., 2005. fMRI reveals two distinct cerebral networks subserving speech motor control. Neurology 64, 700-706.

Rizzolatti, G., Fogassi, L., Gallese, V., 1997. Parietal cortex: from sight to action. Curr.Opin.Neurobiol. 7, 562-567.

Ruff, S., Marie, N., Celsis, P., Cardebat, D., Demonet, J.F., 2003. Neural substrates of impaired categorical perception of phonemes in adult dyslexics: an fMRI study. Brain Cogn 53, 331-334.

Schmidt C.F., Zaehle, T., Meyer, M., Geiser, E., Boesiger, P., Jancke, L., 2007. Silent and continuous fMRI scanningdifferentially modulate activation in an auditory language comprehension task. Human Brain Mapping accepted.

Schubotz, R.I., Friederici, A.D., von Cramon, D.Y., 2000. Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. Neuroimage 11, 1-12.

Segui, J., Dupoux, E., Mehler, J., 1990. The role of the syllable in speech segmentation, phoneme identiication and lexical access. In: Altmann, G., Sillcock, R. (Eds.), Cognitive Models of Speech Processing. MIT Press, Cambridge, pp. 263-280.

Steinhauer, K., Alter, K., Friederici, A.D., 1999b. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat.Neurosci. 2, 191-196.

Steinhauer, K., Alter, K., Friederici, A.D., 1999a. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat.Neurosci. 2, 191-196.

Tervaniemi, M., Hugdahl, K., 2003d. Lateralization of auditory-cortex functions. Brain Res.Brain Res.Rev. 43, 231-246.

Tervaniemi, M., Hugdahl, K., 2003c. Lateralization of auditory-cortex functions. Brain Res.Brain Res.Rev. 43, 231-246.

Tervaniemi, M., Hugdahl, K., 2003b. Lateralization of auditory-cortex functions. Brain Res.Brain Res.Rev. 43, 231-246.

Tervaniemi, M., Hugdahl, K., 2003a. Lateralization of auditory-cortex functions. Brain Res.Brain Res.Rev. 43, 231-246.

Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., Mehler, J., 2005. The role of speech rhythm in language discrimination: further tests with a non-human primate. Developmental Science 8, 26-35.

Van Lancker Sidtis D., Pachana, N., Cummings, J.L., Sidtis, J.J., 2006. Dysprosodic speech following basal ganglia insult: toward a conceptual framework for the study of the cerebral representation of prosody. Brain Lang 97, 135-153.

van Wassenhove, V., Grant, K.W., Poeppel, D., 2007. Temporal window of integration in auditory-visual speech perception. Neuropsychologia 45, 598-607.

Völtz, M., 1991. Das Rhythmusphänomen. Zeitschrift für Sprachwissenschaft 10, 284-296.

Warren, J.E., Wise, R.J., Warren, J.D., 2005. Sounds do-able: auditory-motor transformations and the posterior temporal plane. Trends Neurosci. 28, 636-643.

Wildgruber, D., Ackermann, H., Kreifelts, B., Ethofer, T., 2006. Cerebral processing of linguistic and emotional prosody: fMRI studies. Prog.Brain Res. 156, 249-268.

Wong, P.C., Parsons, L.M., Martinez, M., Diehl, R.L., 2004. The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. J.Neurosci. 24, 9153-9160.

Worsley, K.P., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant signals in images of cerebral activation . Hum.Brain Mapp. 4, 58-73.

Zarate, J.M., Zatorre, R.J., 2005. Neural substrates governing audiovocal integration for vocal pitch regulation in singing. Ann.N.Y.Acad.Sci. 1060, 404-408.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb.Cortex 11, 946-953.

4.3 Study F:

Evidence for a segmental processing device in the human auditory dorsal stream

TINO ZAEHLE [1], EVELINE GEISER [1], KAI ALTER [3], LUTZ JANCKE [1]
& MARTIN MEYER [2]

[1]Departement of Neuropsychology, University of Zurich, Switzerland
[2]Institute of Neuroradiology, University Hospital of Zurich, Switzerland
[3]School of Neurology, Neurobiology & Psychiatry, University of Newcastle, United Kingdome

Abstract

In the present study we investigated the functional organization of sub-lexical auditory perception with specific respect to auditory spectro-temporal processing in speech and nonspeech sounds. Participants discriminated verbal and nonverbal auditory stimuli based on ether spectral or temporal acoustic features in the context of a sparse event-related functional magnetic resonance imaging (fMRI) study.

Based on recent models of language processing, we hypothesized that auditory segmental processing, as required in speech as well as non-speech sound discrimination based on temporal features will lead to a specific involvement of a left-hemispheric dorsal processing network comprising the posterior portion of the inferior frontal cortex and the inferior parietal lobe.

In agreement with our hypothesis results revealed significant responses in the parietal operculum and the inferior frontal gyrus of the left hemisphere when participants had to discriminate speech and nonspeech stimuli based on subtle temporal acoustic features. In contrast, when participants had to discriminate speech and nonspeech stimuli on the base of changes in the frequency content, we explored bilateral activations along the middle temporal gyrus und superior temporal sulcus.

The present data further specify the functioning of a left hemisphere dorsal processing stream in auditory language processing. The results demonstrate a specific involvement of the dorsal pathway in the segmental sub-lexical analysis of speech sounds as well as in the segmental acoustic analysis of non-speech sounds with analogous spectro-temporal characteristics. Thus, the dorsal processing pathway in the auditory domain seems to be specifically dedicated to a general segmental acoustic analysis.

Introduction

The cortical areas devoted to auditory speech processing are located in the middle and superior temporal gyri, and inferior frontal cortical areas of the left and right hemisphere. Since the seminal studies of Fechner and Wernicke (Wernicke, 1874) in the late 19th century, the complementary roles of the left and right temporal lobes in auditory processing have traditionally been examined in brain-lesioned patients with behavioural assessment tools and, in the past few years with functional and structural brain imaging methods. The majority of these studies have supported the predominant role of the left hemisphere in speech processing and point to a stronger involvement of the right hemisphere in music processing (REF). In more recent times attempts have made to identify the properties that might account for the hemispheric differences in auditory processing. Speech related left hemisphere lateralization has been attributed to the linguistic relevance and intelligibility of the input (Scott et al., 2000). Alternatively, it has been argued that the hemispheric specialization exists as a result of asymmetries in basic auditory processing with the left hemisphere preferentially driven by rapidly changing acoustic cues, namely formant transitions (Efron, 1963; Tallal and Piercy, 1973). More specifically, Zatorre and co-workers (Zatorre et al., 2002) have hypothesized that asymmetries in auditory processing may be considered the developmental outcome of optimizing the processing of acoustic cues, with left auditory cortical areas being better adept at temporal resolution and right auditory cortical areas being more amenable to spectral resolution. Along a similar line of argument Poeppel (Poeppel, 2001; Poeppel, 2003) has developed a framework in which asymmetries with respect to speech perception may be accounted for by hemispheric differences in sampling time: the left auditory areas preferentially extract information from short temporal integration windows (~250 Hz) and the right auditory areas from long integration windows (~25 Hz).

These models of functional hemisphere asymmetry in basic auditory processing have supplied the groundwork for large scale neurofunctional

models of speech processing. Underlying these models is the assumption that there are two separate processing streams in the auditory domain, analogous to the dual stream visual processing theory. The notion of a functional segregation in the auditory system is based on seminal findings from experimental work in the rhesus monkey, demonstrating two parallel processing streams specifically devoted to object recognition and object localization (Rauschecker, 1998b; Rauschecker and Tian, 2000). With respect to human speech perception Scott and Wise (2004) have outlined a functional neuroanatomical framework along these lines. According to them an anterior processing stream, running anteriorlateral to primary auditory cortex, is involved in sound-to-meaning mapping and a posterior stream, located in the temporo-parietal-junction, subserves the mapping of speech sounds to motor representations of articulations (Scott and Wise, 2004). The functional anatomy of language elaborated by Hickok and Poeppel (Hickok and Poeppel, 2004; Hickok and Poeppel, 2007) is also based on a distinction between two processing streams; constituting a functional dorsal-ventral partitioning. The ventral stream, projecting ventro-laterally and involving bilaterally the superior temporal sulcus (STS), the posterior middle temporal gyrus (MTG) and inferior temporal gyrus (ITG) mediates the mapping of sound-to-meaning. The dorsal stream, projecting dorso-posteriorly toward the inferior parietal lobe and frontal regions of the left hemisphere,supports the sound-to-articulatory-based representations. This dorsal pathway is taken to play an important role in sub- lexical speech perception (Hickok and Poeppel, 2000). The authors hypothesized that the performance of sublexical tasks specifically engages the dorsal processing path connecting left inferior frontal and left inferior parietal cortex as part of a network for audio-motor integration. Their reasoning is based on lesion data (Caplan et al., 1995) as well as on neuroimaging studies (Burton et al., 2000) that showed dissociations between auditory comprehension tasks and overt sub-lexical phoneme tasks, the former task being associated with the bilateral temporal cortex and the latter task with the inferior frontal cortex., It has been demonstrated convincingly that the left IFG is specifically engaged in sub-lexical tasks that require explicit segmentation of the auditory signal

(Burton, 2001; Burton and Small, 2006). Furthermore, activation of the left IFG has been associated with the discrimination of dynamic auditory cues in speech (Binder et al., 1997; Poldrack et al., 2001; Zatorre et al., 1992) and non-speech sounds (Fiez et al., 1996; Joanisse and Gati, 2003), indicating a general role of the left IFG in processing of transient acoustic information.

In the present study we investigated the neurofunctional representation of sub-lexical auditory processing by comparing hemodynamic responses during the discrimination of speech and non-speech sounds based on temporal or spectral variations. Based on the models of language processing outlined above, we hypothesized that segmental auditory processing occurring when sub-lexical speech discrimination is based on temporal acoustic features, will specifically involve the left-hemispheric dorsal processing network. Furthermore, we hypothesized that the dorsal processing stream is not driven specifically by speech, but is also dedicated to the discrimination of non-speech sounds based on temporal features.

Material and Methods

Subjects

16 native speakers of Swiss-German (age 22–36, mean: 28.43) participated in this study. After a full explanation of the nature and risks of the study, subjects gave informed consent for the participation according to a protocol approved by the local Ethics Committee. They had no history of any neurological, psychiatric or hearing impairment. All subjects were consistently right-handed according to standard handedness questionnaires (Annett, 1992; Steingrüber and Lienert, 1976). In order to avoid gender effects in brain size and shape as well as in functional language lateralization (Luders et al., 2002), only male subjects were included in the study. All subjects were made familiar with the task prior to scanning.

Stimulation

We used a modified version of the stimulus material used in Zaehle and co-workers (2004). Four sets of auditory stimuli were generated (sampling depth: 16 bits, sampling rate of 44.1 kHz) using the SoundForge 4.5 Software (Sonic Foundry Inc., http://www.sonicfoundry.com). Two of these four sets of stimuli were designed to tap the processing of temporal and spectral information without phonological cues (non-speech condition) and two sets with such cues (speech condition). For the non-speech condition, we created Gap-stimuli composed of two sound elements separated by a gap. The leading element was a wideband noise burst with a length of 7 ms. The trailing element was a bandpassed noise (width of 500 Hz) with a duration of 300 ms. All noise bursts were shaped with 0.5-ms linear rise–fall times. Temporal information was controlled by varying the gap durations, resulting in different noise-onset-times (*NOT*) (5, 60, 80, 100 ms). Spectral information was controlled by varying the centre frequencies of the trailing element (1000, 1500, 1600, 1700 HZ) with a bandwidth of +/- 250 Hz.

In the speech condition we used recorded consonant-vowel (CV) syllables, spoken by a trained phonetician. To modify the temporal information different Voice Onset Times (*VOT*) were used, resulting in the perception of different consonants (VOTs in ms /da/=05, /ta/= 60). Variations in spectral information were achieved by modifying the vowel formants of the CV syllables such that they could be perceived as male or female voices. The duration of the syllables ranged from 257 to 269 ms. Figure 1 shows example wave-forms of stimulus pairs

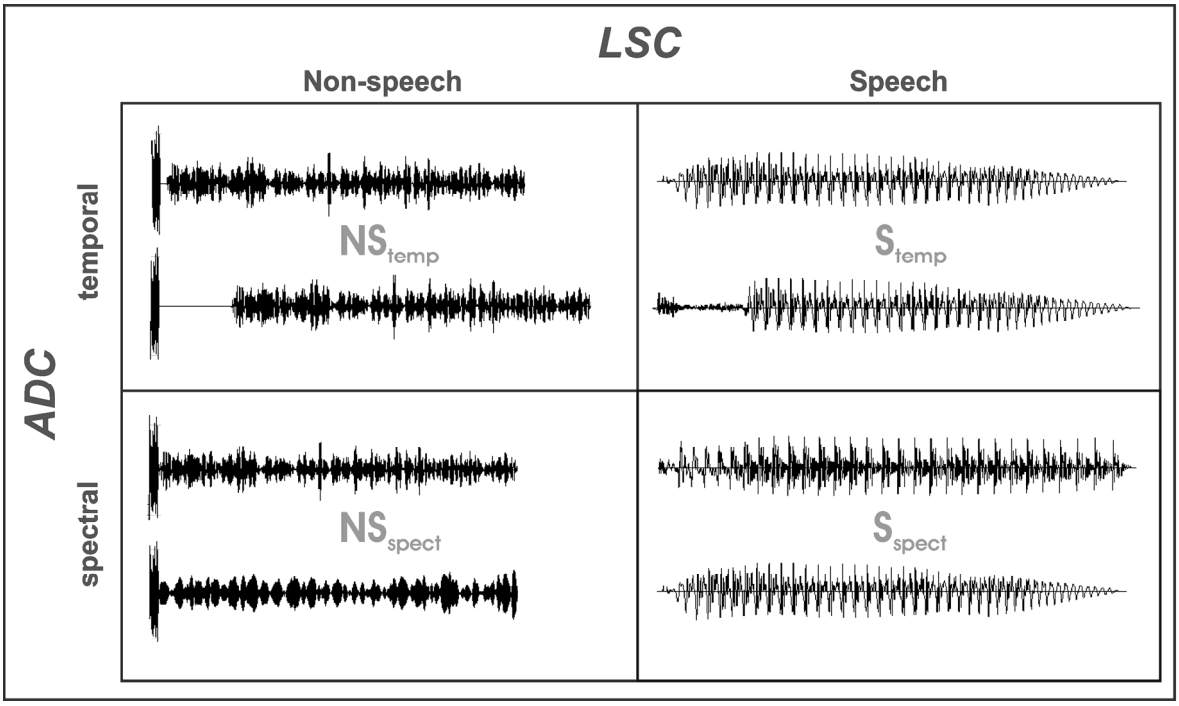

Figure 1: **Figure shows wave-forms of examples of stimuli pairs illustrating the 2x2 design with the factors *linguistic stimulus category (LSC) and acoustic discrimination criterion (ACD)*:**
Left column shows a pair of stimuli used for non-speech sound discrimination based on temporal features (**NS_{temp}**) with varying silent gaps, and a pair of stimuli used for non-speech sound discrimination based on spectral features (**NS_{spect}**) varying centre frequencies of the trailing element. The right column shows a pair of stimuli used for speech sound discrimination based on temporal features (**S_{temp}**) with varying VOTs, and a pair of stimuli used for speech sound discrimination based on spectral features (**S_{spect}**) with varying frequencies contents.

Experimental Design

While lying in the fMRI scanner, participants performed a same-different discrimination task on pairs of speech and non-speech stimuli. Participants had to base their discrimination on temporal or spectral features of the sounds and respond by pressing one of two buttons. Within one trial, participants heard and discriminated five stimulus pairs, all belonging to one of the four experimental conditions (speech or non-speech, temporal or spectral). Figure 2 shows the timing of an individual trial. For each of the four experimental conditions (discrimination of nonspeech sounds based on temporal features **NS_{temp}**, discrimination of nonspeech sounds based on spectral features of **NS_{spect}**, discrimination of speech sounds based on temporal features **S_{temp}**, discrimination of speech sounds

based on spectral features $\mathbf{S_{spect}}$), 20 trials were presented in randomized order, separated into two runs. Twenty one trials of a resting silence condition were included as baseline in each run. During scanning, participants viewed a fixation cross on a projection screen via a mirror fixating their eyes on a cross on a screen. Binaural auditory stimulation was presented by a digital playback system including a high frequency shielded transducer system.

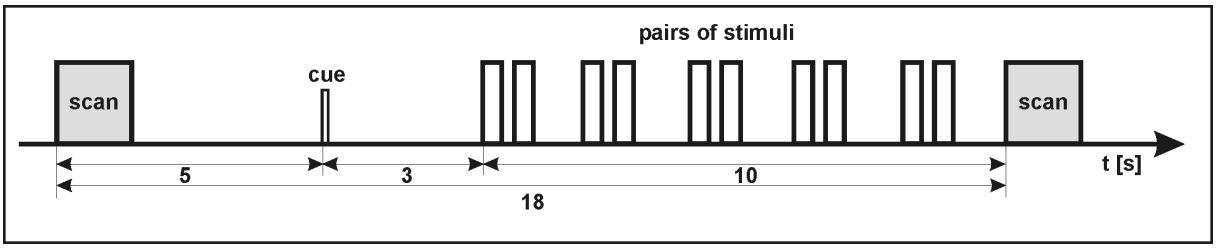

Figure 2: Schematic description of behavioural task:
The stimulation started with a visual cue, a coloured cross which signalled the condition followed 3s later by the first pair of stimuli consisting of two sounds separated by 500 ms silence. The second pair of stimuli followed 2 s after onset of the first pair and so on. In sum 5 pairs of stimuli of the same condition were presented within one trial. The task was to compare the two stimuli of one pair and press on of two response keys to indicate that the sounds were the same or different immediately after the presentation of each pair. 2s after presentation of the last stimulus pair the acquisition of the one fMRI image started.

Data Acquisition

Images were acquired using a 3-T whole body scanner (Philips Intera) equipped with an eight-channel Philips head coil, radiofrequency transmission and signal reception. Using a midsagittal scout image, 16 axial slices (slice thickness 5 mm) of a T2* weighed gradient echo EPI sequence were acquired parallel to the bicommissural plane and covering the whole brain. The following acquisition parameters were used: repetition time (TR):18 s, echo time (TE): 35 ms, flip angle: 90°, voxel size: 1.72 x 1.72 x 5 mm3). Data acquisition was clustered in the first 1.5 seconds of the TR, leaving 16.5 seconds for auditory stimulation without gradient noise (Hall et al. 1999), to use the advantages of a silent data acquisition in auditory fMRI designs. (Gaab, Gabrieli, and Glover 2006a; Gaab, Gabrieli, and Glover 2006b; Schmidt, Zaehle, Meyer, Geiser, Boesiger, and Jancke 2006).

Additionally a high resolution anatomical scans was acquired for each participant.

Data Analysis

Artefact elimination and image analysis was performed using MATLAB 7 (Mathworks Inc., Natiek, MA, USA), the SPM99 software package (Institute of Neurology, London, UK), and the MarsBar toolbox for SPM (http://marsbar.sourceforge.net/) (Brett et al., 2002). Preprocessing of functional MR images included movement correction and coregistration with the anatomical data. The scans were then normalized to a standardized brain space (defined by the Montreal Neurological Institute) using non-linear transformations transformation matrix (voxel size 2x2x2 mm) and smoothed using a Gaussian kernel with 6 mm fullwidth half-maximum (FWHM). The EPI time series was analyzed using a general linear model analysis (Friston et al., 1995). Separate regressors were defined for each condition using a box-car function without convolving with the canonical haemodynamic response function as reference waveform. After estimation of model parameters for each subject, an analysis of variance was calculated for the whole group, using the individual contrast images for the main effects. In a first step, we compared the four conditions separately with the resting baseline. Results of these statistics were thresholded by T = 7.7 (p = 0.05 corrected for multiple comparisons). In order to quantify functional hemisphere differences of local brain activity in auditory fields stretching along the supratemporal plane (STP) as a function of auditory processing, subsequently region of interest (ROI) analyses were conducted. For that purpose a set of six distinct regions of interest (ROI) was placed for each subject in Heschl's gyrus (HG, middle STP), the planum polare (PP, anterior STP) and planum temporale (PT, posterior STP) of each hemisphere. ROIs were anatomically defined based on macroanatomical landmarks (Rademacher, et al. 2001; Steinmetz, et al. 1990). Mean beta-values were collected for each subject from each ROI (HG, PP, PT) and subjected to a repeated-measure within-subjects (2 x 2 x 2) ANOVA with the factors

linguistic stimulus category (LSC) (Speech-Nonspeech), *acoustic discrimination criterion* (*ADC)* (Temporal-Spectral), and *hemisphere* (left, right).

In a next step subtractions between conditions were performed. The two principle results discussed are those that evaluate the statistical significance of the two main effects. The effect of *acoustic discrimination criterion (ADC)* was analysed by comparing the processing of both, speech and non-speech sounds based on temporal with the processing of these sounds based on spectral variations. The effect of *linguistic stimulus category (LSC)* was evaluated by comparing the processing of the speech sounds with the processing of the non-speech sounds independent of the temporal or spectral variations. Results of these statistics were thresholded by T = 3.73 (p = 0.001 uncorrected for multiple comparisons) and k = 30 voxels. Additional, to further explore the statistical results of these main effects, functionally defined ROIs were examined. Here we used the peak activated cortical voxel as the centre and a sphere with a radius of 6 mm. To test for main effects of *ADC* and *LSC* a 2x2 ANOVAs with the factor *LSC* (Speech-Nonspeech) and *ADC* (Temporal-Spectral) was run for each of these ROIs.

Results

Behavioural Results

Behavioural data were successfully collected from all participants. Mean accuracy data (percent correct responses) were subjected into a 2 x 2 ANOVA with factors *LSC* (Speech-Nonspeech) and *ADC* (Temporal-Spectral). Analysis revealed no significant main effects of the factor *LSC* ($F(1,15) = 0.1$, $p = 0.9$) nor of the factor *ADC* ($F(1,15) = 0.2$, $p = 0.7$), demonstrating that the mean accuracies did not differ between the speech and non-speech condition, nor between the spectral and temporal discrimination. Overall, the mean accuracy with which the subjects performed the tasks was 86% of trials.

Functional MRI Results

In a first step, each of the four stimulus conditions was compared with the resting baseline condition. As illustrated in Figure 3, analysis revealed significant activation along the bilateral superior temporal plane for all four conditions. Subsequent ROI analysis on anatomically predefined regions (PP, HG, PT) showed a general functional leftward asymmetry for all conditions and regions (cf. Figure 4). Separate 2x2x2 ANOVAs with the factor LSC (speech/nonspeech), ADC (temporal/spectral), and hemisphere (left/right) revealed a significant main effect of the factor hemisphere for PP ($F(1,15) = 11.12$, $p = 0.005$), PT ($F(1,15) = 22.83$, $p = 0.000$), and HG ($F(1,15) = 7.15$, $p = 0.017$), demonstrating stronger responses in the left as compared to the right hemisphere. The HG also explored a significant main effect of the factor LSC ($F(1,15) = 7.47$, $p = 0.015$), indicating stronger responses during the speech condition, and a significant main effect of the factor ADC ($F(1,15) = 6.17$, $p = 0.025$), showing stronger responses for the spectral conditions.

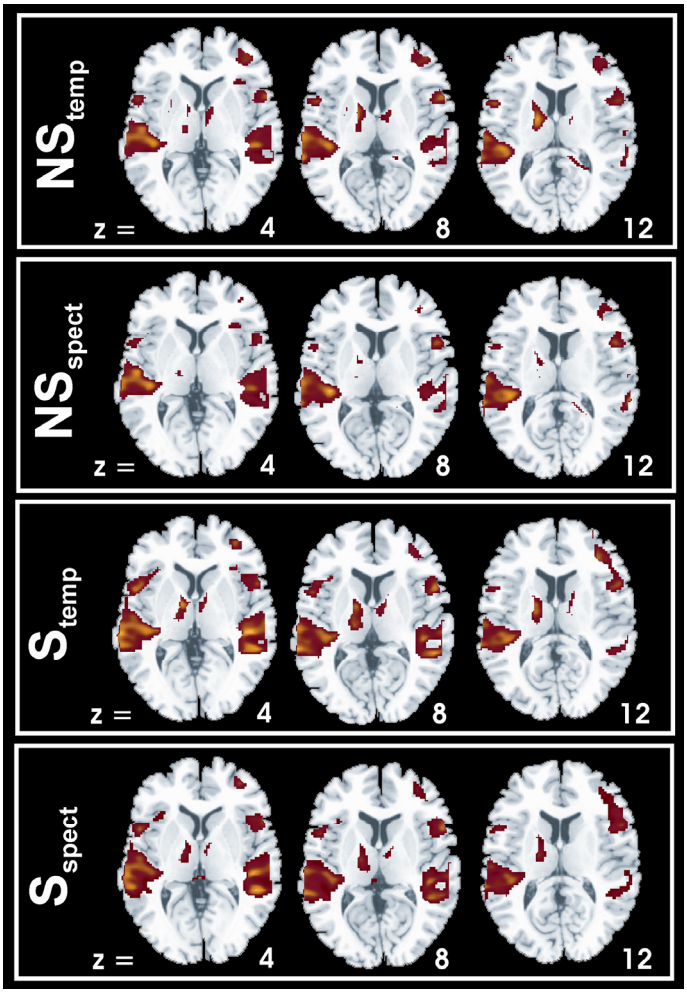

Figure 3: Activation pattern for the comparison for each condition with the resting baseline. Results are superimposed onto horizontal sections of a standard anatomical template. Data is plotted in neurological convention.

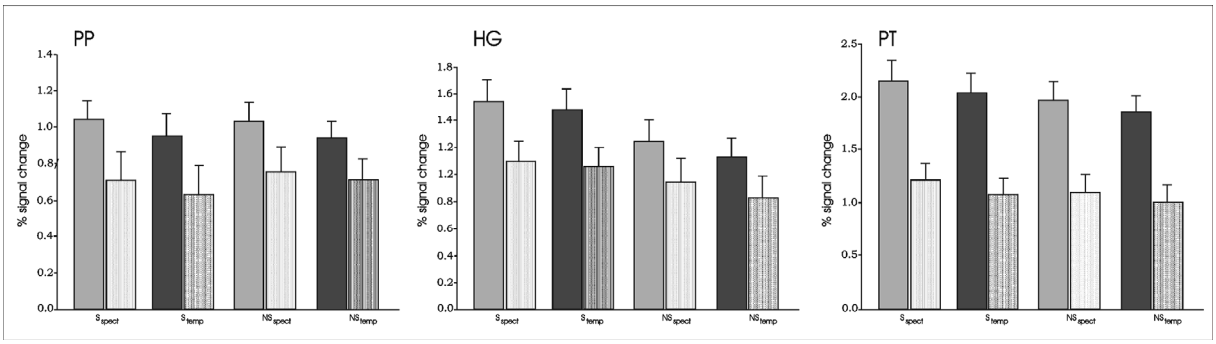

Figure 4: **Analysis of three bilateral superior temporal plane regions** for the main effects of *LSC* (speech/nonspeech), *ACD* (temporal/spectral) and *hemisphere* (left/right). Mean *beta*-values averaged across all subjects within bilateral planum polare (PP), Heschl's gyrus (HG), and planum temporale (PT). Error bars indicate the standard deviation. Light gray bars refer to the temporal conditions; dark gray bars refer to the spectral condition. Bars in plain-colour refer to the left; bars in patterned colour refer to the right hemisphere.

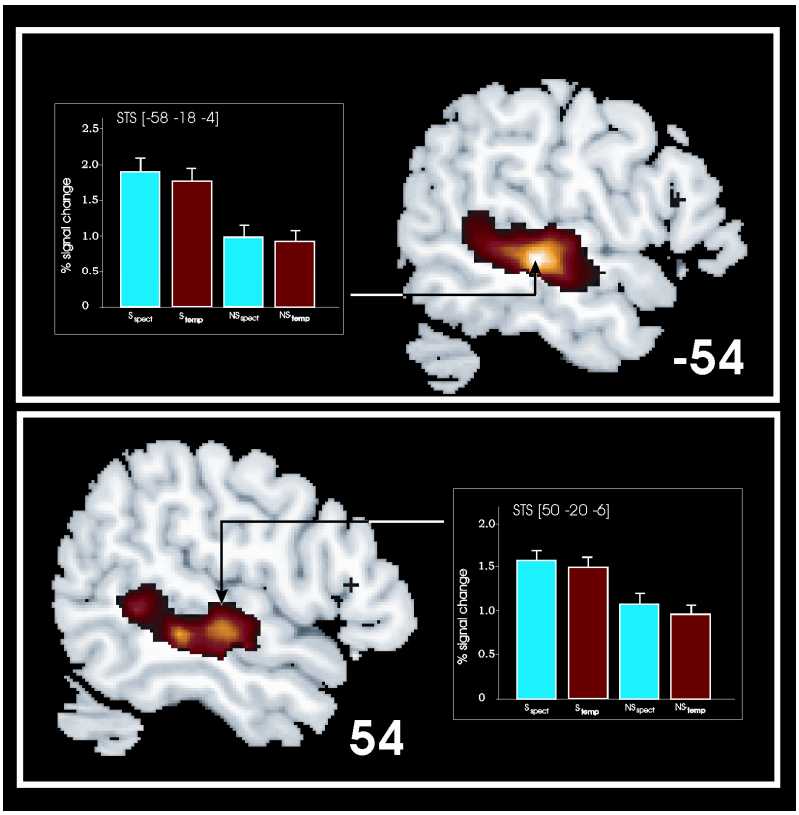

Figure 5: Group main effects for the factor *LSC*:
Stronger responses to Speech sounds are colour-coded at a red-yellow scale. Comparison of Non-speech sounds in relation to Speech sounds did not reveal significant responses. Results are superimposed onto sagital sections of a standard anatomical template (x = -54, 54). Data is plotted in neurological convention. Bar plots show mean signal increase for the four conditions (discrimination of nonspeech sounds based on temporal features $\mathbf{NS_{temp}}$, discrimination of nonspeech sounds based on spectral features of $\mathbf{NS_{spect}}$, discrimination of speech sounds based on temporal features $\mathbf{S_{temp}}$, discrimination of speech sounds based on spectral features $\mathbf{S_{spect}}$) in particular regions (right superior temporal sulcus STS [50 -20 -6], and left superior temporal sulcus STS [-58 -18 -4], exploring stronger responses to speech then to nonspeech sounds)

Main effect of ***LSC***

The speech condition as compared to the nonspeech condition activated the auditory cortices in the superior temporal sulcus (STS) and middle temporal gyrus (MTG) bilaterally regardless of the acoustic variations. Additionally, frontal activation was observed for the processing of speech sounds in the left middle frontal gyrus (MFG) (cf. Figure 5 and Table 1). The comparison of the nonspeech condition with the speech condition did not reveal significant activation. Figure 5 also shows the mean percent signal change for two regions of interest (left and right STS). Results of a subsequent 2x2 ANOVAs with the factors *LSC* (speech/nonspeech), and *ADC* (temp/spect) separately for each ROI are summarized in Table 3. All regions display a main effect of the factor *LSC*, indicating significantly stronger hemodynamic responses to speech than to nonspeech stimuli, and a significant effect of the factor *ADC*; reflecting stronger responses to spectral than temporal variations in both the speech and nonspeech condition.

Table 1: fMRI main effects of factor linguistic stimulus category (LSC)

Condition and anatomical area	L/R	*T*-value	x	y	z
Speech > Non-speech					
Superior temporal sulcus	L	14.33	-58	-18	-4
	R	9.70	50	-20	-6
Middle frontal gyrus	L	4.60	-40	8	38
Cingulat gyrus	L	7.10	-4	-20	30
	R	4.38	8	-58	38
Thalamus	L	5.83	-12	-14	10
Non-speech > Speech					
NO suprathreshold voxels					

Main effect of ***ADC***

As shown in Figure 6 and listed in Table 3, the temporal condition in comparison to the spectral condition activated the MFG and the frontal operculum (FO) bilaterally, and the parietal operculum (PO), and the inferior frontal gyrus (IFG) of the left hemisphere. Subsequent ROI analyses further corroborate these results (cf. Table 1). All regions displayed a significant main effect of the factor *ADC*, and no main effect of the factor *LSC*, demonstrating stronger hemodynamic responses during the discrimination task based on temporal features as compared to the discrimination based on spectral features independent of the "speechness" of the sounds.

In contrast, comparison of the spectral condition with the temporal condition revealed activations of the anterior and middle parts of the STS, and the Middle occipital gyrus of both hemispheres (cf. Figure 6 and Table 3). As illustrated in Figure 6 and listed in Table 1, analysis of ROI data revealed significant main effect of the factor *ADC*, demonstrating significant stronger hemodynamic responses during the categorization based on spectral features as compared to the categorization based on the temporal features in all regions of interest. Additionally, the left middle STS, the right middle STS, and the right anterior STS also showed a main effect of the factor *LSC*, indicating stronger responses to speech then non-speech sounds.

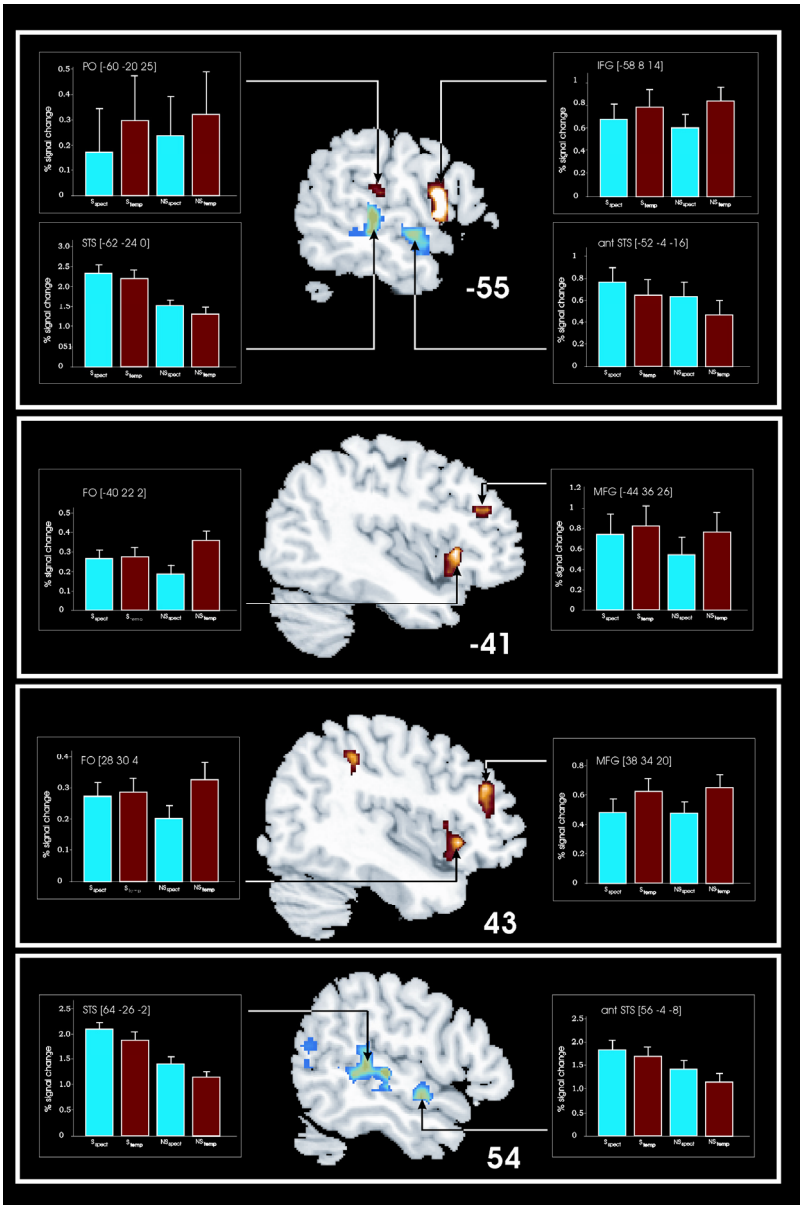

Figure 6: Group main effects for the factor *ACD:*
Stronger responses to temporal processing are colour-coded at a red-yellow scale. Starker responses to spectral processing are colour-coded at a blue-petrol scale. Results are superimposed onto sagital sections of a standard anatomical template (x = -55, -41, 43, 54). Data is plotted in neurological convention. Bar plots show mean signal increase for the four conditions (discrimination of nonspeech sounds based on temporal features $\mathbf{NS_{temp}}$, discrimination of nonspeech sounds based on spectral features of $\mathbf{NS_{spect}}$, discrimination of speech sounds based on temporal features $\mathbf{S_{temp}}$, discrimination of speech sounds based on spectral features $\mathbf{S_{spect}}$) in particular regions (right middle frontal gyrus MFG [38 34 20], left middle frontal gyrus MFG [-44 36 26], right frontal operculum FO [28 30 -4], left frontal operculum FO [-40 22 2], left inferior frontal gyrus IFG [-58 8 14], and left parietal operculum PO [-60 -16 28] exploring starker responses for the temporal then spectral modulations, and right anterior medial temporal gyrus MTG [56 -4 -8], left anterior medial temporal gyrus MTG [-52 -4 -16], right medial temporal gyrus MTG [64 -26 -2], left medial temporal gyrus MTG [-62 -24 0] exploring a stronger response to spectral modulations).

Table 2: Main effects of the factor acoustic

Condition and anatomical area	L/R	T-value	x	y	z
Temp > Spect					
Inferior frontal gyrus	L	7.25	-58	8	14
Pariatal operculum	L	4.89	-60	-20	25
Middle frontal gyrus	L	4.37	-44	36	26
	R	6.06	38	34	20
	R	5.44	34	44	8
Frontal operculum	L	5.62	-40	22	2
	R	6.62	28	30	-4
Nucleus caudatus	R	5.87	10	4	18
Spect > Temp					
Superior temporal sulcus	L	9.35	-62	-24	0
	R	9.11	64	-26	-2
ant. Superior temporal sulcus	L	5.25	-52	-4	-16
	R	5.33	56	-4	-8
Middle occipital gyrus	L	4.25	-40	-78	20
	R	6.17	46	-72	20
Cuneus / Precuneus	L	4.42	-12	-80	40

Discussion

The aim of the present study was to explore the functional organisation of sub-lexical auditory perception with respect to the spectro-temporal processing in speech and nonspeech sounds. Results show that the discrimination of auditory sounds, when separately compared to a silent baseline, leads to significant activation of the bilateral superior temporal plane with a clear functional leftward asymmetry. In former studies we and others have shown that areas along the superior temporal plane, namely the HG and PT, are specifically dedicated to the processing of rapidly changing acoustic cues in speech and non-speech sounds (Binder et al., 2000; Jancke et al., 2002; Joanisse and Gati, 2003; Meyer et al., 2005; Zaehle et al., 2004).

Table 3: Summary of results of the 2x2 ANOVAs carried out to test the effects of LSC and ACD on the hemodynamic responses in particular regions.

	Factor				Interaction	
	LSC **(df: 1.15)**		***ACD*** **(df: 1.15)**		***LSC x ACD*** **(df:1.15)**	
ROI	**F**	**p**	**F**	**P**	**F**	**p**
Speech > Non-speech						
STS [-58 -18 -4]	160.4	0.000	5.69	0.031	0.532	0.477
STS [50 -20 -6]	77.02	0.000	10.9	0.006	0.139	0.715
MFG [-40 8 38]	20.26	0.000	3.135	0.096	0.967	0.341
Temp > Spect						
IFG [-58 8 14]	0.14	0.9	69.844	0.000	1.563	0.23
PO [-60 -20 25]	0.52	0.48	58.65	0.000	0.861	0.368
FO [-40 22 2]	0.17	0.68	11.31	0.004	9.60	0.007
FO [28 30 -4]	0.004	0.95	30.47	0.000	7.59	0.015
MFG [-44 36 26]	1.68	0.21	13.93	0.002	1.339	0.265
MFG [38 34 20]	0.03	0.86	30.08	0.000	0.065	0.802
Spect > Temp						
STS [-62 -24 0]	107.5	0.000	78.62	0.000	0.549	0.47
STS [64 -26 -2]	25.38	0.000	17.08	0.000	0.215	0.649
ant. STS [-52 -4 -16]	2.63	0.126	23.26	0.000	0.868	0.366
ant. STS [56 -4 -8]	10.25	0.006	24.19	0.000	1.013	0.33

In the present study we confirmed these former results by demonstrating consistent left hemispheric functional asymmetry for all conditions regardless of the *ADC* or the "speechness" of the sounds. Since all of the stimuli contained rapidly changing spectral information, the data lead to the assumption that the superior temporal plane is involved in the fine grained analysis of spectro-temporal characteristic of sounds in general.

Effects of LSC

When comparing the processing of speech and nonspeech sounds stronger responses were found bilaterally in the middle portion of the superior temporal region with a subdominace of the left hemisphere during the processing of speech sounds. The middle portion of the STS and adjacent areas in the STG bilaterally have consistently been implicated in the phonological analysis of speech sounds (Binder et al., 2000; Indefrey and Levelt, 2004; Liebenthal et al., 2005; Mummery et al., 1999; Price et al., 1996; Uppenkamp et al., 2006; Zatorre et al., 1992). As suggested by Hickok and Poeppel (2007), portions of the bilateral STS are driven by acoustic signals that contain phonological information. Our results are in agreement with these former findings and provide further evidence for the eminent role of the bilateral STS in phonological-level processes during speech perception.

Effects of ADC

The comparison of the tasks in which participants had to discriminate non-speech and speech stimuli by changes in spectral properties in relation to changes in temporal modifications, showed stronger activations of the bilateral STS for spectral based discrimination. To accomplish this task, subjects had to detect and compare differences in the spectral characteristics of the auditory stimuli. There is a growing body of evidence showing that the processing of spectral properties, as characteristic features of auditory sounds, is associated with activations of the STS in both

hemispheres. In particular, it has been demonstrated that bilateral STS activations are associated with the processing of spectral features of auditory events as required during the perception of pitch (Griffiths et al., 1998), melody (Patterson et al., 2002), harmonic tones (Hall et al., 2002), vowels (Obleser et al., 2006; Uppenkamp et al., 2006), and voice (Kriegstein and Giraud, 2004; Warren et al., 2006). Furthermore, bilateral STS/STG activations have been reported for the processing of spectrally relevant acoustic properties compared to temporally relevant acoustic properties in non-speech sounds using PET (Zatorre and Belin, 2001) and fMRI (Jamison et al., 2005). Our results are in line with these findings and further support the notion of an anterolaterally directed processing stream specialized for the processing of auditory sound features (Rauschecker, 1998a; Rauschecker and Tian, 2000). The STS/STG region has been taken to constitute part of the "what" stream that is specialized for auditory-feature processing, spectral information playing an important role. However, the posterior portion of the bilateral STS activation in response to spectral discrimination also showed stronger activity for the speech as compared to the non-speech sounds. It has been argued that the "what" stream is involved in perceiving auditory objects by extracting the invariant set of auditory features that allow for the identification of sound sources and thus may contribute to the identification of the speaker (Belin and Zatorre, 2000). In the context of the present study, the activation of the bilateral STS in the spectral condition is associated with the processing of spectral features of sounds as the relevant "acoustic invariants" on which task performance resides. When participants had to discriminate the non-speech and speech stimuli on the basis of temporal variations, a task that requires a segmental analysis of the auditory stimuli, we identified a network comprising the frontal operculum (FO) and middle frontal gyri bilaterally and the parietal operculum (PO) and the posterior part of the inferior frontal gyrus (IFG) of the left hemisphere. In general, the left frontal cortex plays an important role in phonological processing. As has been shown the categorical perception of VOT is impaired in non-fluent aphasic patients with left frontal cortex lesions (Blumstein et al., 1977a; Blumstein et al., 1977b). Japanese speaking individuals with lesions of the

left IFG show impairments in processing phonologically coded Kana while having no difficulties processing orthographically coded Kanji (Sasanuma and Monoi, 1975). Furthermore, several neuroimaging studies demonstrated a specific involvement of the inferior and middle frontal gyri in the processing of phonemes (Burton et al., 2000; Joanisse and Gati, 2003) and sounds with fast acoustic transients (Joanisse and Gati, 2003; Johnsrude et al., 1997; Poldrack et al., 2001), suggesting specific sensitivity of the left IFG for the processing of speech and non-speech sounds that contain rapidly changing information at a time range similar to speech. It has been argued that the process of segmenting the initial consonant from the following vowel in speech discrimination may be viewed as the conceptual correlate of inferior frontal activation (Burton, 2001; Burton et al., 2000). The present study confirms the specific involvement of the left IFG in discriminating speech sounds on the basis of temporal but not spectral feature variations. Thus the left IFG seems to be specifically dedicated to a segmental analysis of sub-lexical speech sounds. Furthermore, by obtaining the same pattern of left IFG activation during the discrimination of non-speech sounds on the basis of temporal but not spectral features we provide evidence that the left IFG functioning is not associated with the linguistic nature of the processed auditory sounds.

A further important observation of the present study is the specific activation for the discrimination of speech and non-speech sounds based on temporal features of an area in the convex of the anterior part of the supramarginal gyrus and the inferior postcentral gyrus constituting the parietal operculum (PO). It has been demonstrated cytoarchitectonically that the PO contains auditory-related cortex and plays a role in auditory function (Galaburda and Sanides, 1980). Furthermore, electrophysiological measurements of the exposed cortex showed that the PO explores auditory evoked potentials similar to that of the secondary auditory area of the lower lip of the sylvian fissure (Celesia, 1976). The PO also displays a left-right asymmetry in the surface pattern (Steinmetz et al., 1990), and is larger in the left compared to the right hemisphere in the most right handed subjects

(Habib et al., 1995). It has been shown that a reversed (rightward) PO asymmetry forms the best prediction of non-right handedness irrespective of the pattern of PT asymmetry (Habib et al., 1995), and leftward PO asymmetry has been related to left hemisphere language lateralization. Furthermore, morphological studies demonstrated that adults with developmental dyslexia exhibit a stronger left-sided asymmetry of the PO as compared to normal controls (Robichon et al., 2000), and that dyslexic adults as compared to normal controls show decreases in white matter of the PO region (Silani et al., 2005). Habib (2000) suggested that parietal rather than temporal asymmetries may be the most relevant morphological characteristic of the dyslexic brain (Habib, 2000). Furthermore, a recent fMRI investigation showed activation of the left inferior parietal cortex specifically linked to phonological processes in controls (Dufor et al., 2007). This study also showed that individuals with developmental dyslexia fail to specifically recruit the left PO during phoneme categorization. Additionally, lesion studies identified the left PO as a principal site of phonemic processing in speech perception (Caplan et al., 1995). In the present study we demonstrate that the left PO is specifically involved in the discrimination of speech and non-speech sounds based on temporal features but not by means of the spectral features of the sounds.

In essence, the results underscore a particular preference of the left dorsal stream comprising the posterior portion of the inferior frontal cortex and the parietal operculum for the processing of temporal features available in the acoustic signal.

Recent brain imaging studies have been providing corroborating evidence for a hemispheric specialization within the auditory cortex by demonstrating preference of left auditory cortices for rapidly changing acoustic cues available in speech and nonspeech sounds (Jancke et al., 2002; Zaehle et al., 2004) and a dominance of right perisylvian cortex for the processing of spectral profiles (Boemio et al., 2005; Schonwiesner et al., 2005). Our data are in line with these observations in that we explored a

dominance of the left hemispheric auditory cortex in the initial acoustic analysis of speech- as well as non-speech sounds with analogous spectro-temporal features. As proposed by Hickok and Poeppel (2007) the auditory signal passes through several stages of analysis. The earliest stage carries out a spectro-temporal analysis and appears in the bilateral, but left-subdominant supratemporal plane. The present study confirms this assumption and further evidences the importance of the left HG and left PT in the acoustic analysis of auditory signals independent of the linguistic content. Subsequently after the initial spectro-temporal analysis the processing system diverges into a ventral pathway and a dorsal pathway. The ventral pathway is supposed to map sensory or phonological representations onto lexical conceptual representations. In the present study we showed a specific involvement of the middle STS for the processing of consonant-vowel-syllables and thus further evidence the elevated role of the bilateral STS in phonological-level processes during speech perception. The dorsal processing stream is supposed to be specifically involved in speech perception tasks like syllable discrimination. Our results clearly demonstrate an involvement of a dorsal processing stream, in particular of the left posterior IFG and left PO during the discrimination of speech and non-speech sounds by means of temporal features. Thus we suggest that the functioning of the dorsal stream in auditory speech processing is dedicated to auditory segmentation and is independent from the linguistic content of the processed auditory input.

It has been suggested, that the dorsal fronto-parietal network predominantly in the left hemisphere might be tuned in ontogenesis, because attention to phonetic segments provides the base for the learning process of articulating speech and is functioning to interface auditory and articulatory representations of speech (Hickok and Poeppel, 2000; Hickok and Poeppel, 2004; Hickok and Poeppel, 2007). This functioning will be continued to function in adults providing the basic for acquiring new vocabulary. Interestingly, several studies demonstrated malfunctioning of the left IFG (Temple et al., 2000) and the left PO (Dufor et al., 2007) in

children and adults with developmental dyslexia during phonological processing (Cao et al., 2006; Hoeft et al., 2007). It has also been reported that in dyslexic adults morphological deviations occur in left prefrontal and left parietal areas in high correlation with deficits in processing rapidly presented sounds (Jernigan et al., 1991). Thus malfunctioning of the left dorsal processing stream during ontogenesis might be a general neurofunctional base for the development of language related impairments like dyslexia.

In sum, our data show that at an initial stage spectro-temporal analysis of speech and non-speech sounds is performed by bilateral, but left dominant superior temporal cortex. Furthermore we demonstrate that the processing of temporal information in both, speech and non-speech sounds, which might be related to the extraction of segmental information specifically engaged a left hemispheric fronto-parietal cortical network. Thus our fMRI data provide additional evidence for the existence of a dorsal processing stream in auditory speech processing. Moreover, by evidencing that this cortical network is specifically activated when extraction of temporal or segmental information is required to solve an auditory discrimination task, we found new evidence to enhance the present knowledge of dorsal auditory pathway functions and demonstrate that this pathway is specifically dedicated to serve general segmental acoustic analyses of sounds rather then displaying specificity for speech processing.

Acknowledgements

This work was supported by Swiss National Science Foundation Grant No. 46234103 (TZ, EG) and Swiss SNF 46234101 (MM).

Reference List

Annett M. 1992. Five tests of hand skill. Cortex 28:583-600.

Belin P, Zatorre RJ. 2000. 'What', 'where' and 'how' in auditory cortex. Nat.Neurosci. 3:965-966.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb.Cortex 10:512-528.

Binder JR, Fros, JA, Hammeke TA, Cox RW, Rao SM, Prieto T. 1997. Human brain language areas identified by functional magnetic resonance imaging. J.Neurosci. 17:353-362.

Blumstein SE, Baker E, Goodglass H. 1977a. Phonological factors in auditory comprehension in aphasia. Neuropsychologia 15:19-30.

Blumstein SE, Cooper WE, Zurif EG, Caramazza A. 1977b. The perception and production of voice-onset time in aphasia. Neuropsychologia 15:371-383.

Boemio A, Fromm S, Braun A, Poeppel D. 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat.Neurosci. 8:389-395.

Brett M, Anton JL, Valabregue R, Poline JB. 2002. Region of interest analysis using an SPM toolbox. NeuroImage 16 (2) (Abstract; Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2 –6, 2002, Sendai, Japan; http://marsbar.sourceforge.net/)

Burton M. 2001. The role of inferior frontal cortex in phonological processing. Cognitive Science 25:695-709.

Burton MW, Small SL. 2006. Functional neuroanatomy of segmenting speech and nonspeech. Cortex 42:644-651.

Burton MW, Small SL, Blumstein S.E. 2000. The role of segmentation in phonological processing: an fMRI investigation. J.Cogn Neurosci. 12:679-690.

Cao F, Bitan T, Chou TL, Burman DD, Booth JR. 2006. Deficient orthographic and phonological representations in children with dyslexia revealed by brain activation patterns. J.Child Psychol.Psychiatry 47:1041-1050.

Caplan D, Gow D, Makris N. 1995. Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. Neurology 45:293-298.

Celesia GG. 1976. Organization of auditory cortical areas in man. Brain 99:403-414.

Dufor O, Serniclaes W, Sprenger-Charolles L, Demonet JF. 2007. Top-down processes during auditory phoneme categorization in dyslexia: A PET study. Neuroimage 34:1692-1707.

Efron R. 1963. Temporal Perception, Aphasia and D'ej'a vu. Brain 86:403-424.

Fiez JA, Raichle ME, Balota DA, Tallal P, Petersen SE. 1996. PET activation of posterior temporal regions during auditory word presentation and verb generation. Cereb.Cortex 6:1-10.

Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R. 1995. Analysis of fMRI time-series revisited. Neuroimage. 2:45-53.

Gaab N, Gabrieli JDE, Glover GH. 2006a. Assessing the influence of scanner background noise on auditory processing - I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. Hum.Brain.Mapp (DOI: 10.1002/hbm.20298)

Gaab N, Gabrieli JDE, Glover GH. 2006b. Assessing the influence of scanner background noise on auditory processing - II: An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. Hum.Brain.Mapp (DOI: 10.1002/hbm.20299)

Galaburda A, Sanides F. 1980. Cytoarchitectonic organization of the human auditory cortex. J.Comp Neurol. 190:597-610.

Griffiths TD, Buchel C, Frackowiak RS, Patterson RD. 1998. Analysis of temporal structure in sound by the human brain. Nat.Neurosci. 1:422-427.

Habib M. 2000. The neurological basis of developmental dyslexia: an overview and working hypothesis. Brain 123:2373-2399.

Habib M, Robichon F, Levrier O, Khalil R, Salamon G. 1995. Diverging asymmetries of temporo-parietal cortical areas: a reappraisal of Geschwind/Galaburda theory. Brain Lang 48:238-258.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. 1999. "Sparse" temporal sampling in auditory fMRI. Hum.Brain Mapp. 7:213-223.

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ. 2002. Spectral and temporal processing in human auditory cortex. Cereb.Cortex 12:140-149.

Hickok G, Poeppel D. 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci. 4:131-138.

Hickok G, Poeppel D. 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92:67-99.

Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nat.Rev.Neurosci. 8:393-402.

Hoeft F, Meyler A, Hernandez A, Juel C, Taylor-Hill H, Martindale JL, McMillon G, Kolchugina G, Black JM, Faizi A, Deutsch GK, Siok WT, Reiss AL, Whitfield-Gabrieli S, Gabrieli JD. 2007. Functional and morphometric brain dissociation between dyslexia and reading ability. Proc.Natl.Acad.Sci.U.S A 104:4234-4239.

Indefrey P, Levelt WJ. 2004. The spatial and temporal signatures of word production components. Cognition 92:101-144.

Jamison HL, Watkins KE, Bishop DV, Matthews PM. 2005. Hemispheric Specialization for Processing Auditory Nonspeech Stimuli. Cereb.Cortex. (DOI: 10.1093/cercor/bhj068)

Jancke L, Wustenberg T, Scheich H, Heinze HJ. 2002. Phonetic perception and the temporal cortex. Neuroimage. 15:733-746.

Jernigan TL, Hesselink JR, Sowell E, Tallal PA. 1991. Cerebral structure on magnetic resonance imaging in language- and learning-impaired children. Arch.Neurol. 48:539-545.

Joanisse MF, Gati JS. 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. Neuroimage. 19:64-79.

Johnsrude IS, Zatorre RJ, Milner BA, Evans AC. 1997. Left-hemisphere specialization for the processing of acoustic transients. Neuroreport 8:1761-1765.

Kriegstein KV, Giraud AL. 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22:948-955.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. 2005. Neural substrates of phonemic perception. Cereb.Cortex 15:1621-1631.

Luders E, Steinmetz H, Jancke L. 2002. Brain size and grey matter volume in the healthy human brain. Neuroreport 13:2371-2374.

Meyer M, Zaehle T, Gountouna VE, Barron A, Jancke L, Turk A. 2005. Spectro-temporal processing during speech perception involves left posterior auditory cortex. Neuroreport 16:1985-1989.

Mummery CJ, Ashburner J, Scott SK, Wise RJ. 1999. Functional neuroimaging of speech perception in six normal and two aphasic subjects. J.Acoust.Soc.Am. 106:449-457.

Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. 2006. Vowel sound extraction in anterior superior temporal cortex. Hum.Brain Mapp. 27:562-571.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD, 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767-776.

Poeppel D. 2001. Pure word deafness and the bilateral processing of the speech code. Cognitive Science 25:679-693.

Poeppel D. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication 41:245-255.

Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, Gabrieli JD. 2001. Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. J.Cogn Neurosci. 13:687-697.

Price CJ, Wise RJ, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RS, Friston KJ. 1996. Hearing and saying. The functional neuro-anatomy of auditory word processing. Brain 119:919-931.

Rauschecker JP. 1998a. Cortical processing of complex sounds. Curr.Opin.Neurobiol. 8:516-521.

Rauschecker JP. 1998b. Parallel processing in the auditory cortex of primates. Audiol.Neurootol. 3:86-103.

Rauschecker JP, Tian B. 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc.Natl.Acad.Sci.U.S A 97:11800-11806.

Robichon F, Levrier O, Farnarier P, Habib M. 2000. Developmental dyslexia: atypical cortical asymmetries and functional significance. Eur.J.Neurol. 7:35-46.

Sasanuma S, Monoi H. 1975. The syndrome of Gogi (word meaning) aphasia. Selective impairment of kanji processing. Neurology 25:627-632.

Schmidt CF, Zaehle T, Meyer M, Geiser E, Boesiger P, Jancke L. 2007. Silent and continuous fMRI scanning differentially modulate activation in an

auditory language comprehension task. Hum.Brain.Mapp. (DOI: 10.1002/hbm.20372)

Schonwiesner M, Rubsamen R, von Cramon DY. 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur.J.Neurosci. 22:1521-1528.

Scott SK, Blank CC, Rosen S, Wise RJ. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123: 2400-2406.

Scott SK, Wise RJ. 2004. The functional neuroanatomy of prelexical processing in speech perception. Cognition 92:13-45.

Silani G, Frith U, Demonet JF, Fazio F, Perani D, Price C, Frith CD, Paulesu E. 2005. Brain abnormalities underlying altered activation in dyslexia: a voxel based morphometry study. Brain 128:2453-2461.

Steingrüber HJ, Lienert GA. 1976. Hand-Dominanz-Test (HDT). Hogrefe, Göttingen, Toronto, Zürich.

Steinmetz H, Ebeling U, Huang YX, Kahn T. 1990. Sulcus topography of the parietal opercular region: an anatomic and MR study. Brain Lang 38:515-533.

Tallal P, Piercy M. 1973. Defects of non-verbal auditory perception in children with developmental aphasia. Nature 241:468-469.

Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM, Gabrieli JD. 2000. Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. Proc.Natl.Acad.Sci.U.S.A 97:13907-13912.

Uppenkamp S, Johnsrude IS, Norris D, Marslen-Wilson W, Patterson RD. 2006. Locating the initial stages of speech-sound processing in human temporal cortex. Neuroimage 31:1284-1296.

Warren JD, Scott SK, Price CJ, Griffiths TD. 2006. Human brain mechanisms for the early analysis of voices. Neuroimage 31:1389-1397.

Wernicke C. 1874. Der aphasische Symptomkomplex: eine psychologische Studie auf anatomischer Basis. The dual brain: hemispheric specialization in humans. Cohn & Weigert, Breslau.

Zaehle T, Schmidt CF, Meyer M, Baumann S, Baltes C, Boesiger P, Jancke L. (in press). Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. Neuroimgae

Zaehle T, Wustenberg T, Meyer M, Jancke L. 2004. Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur.J.Neurosci. 20:2447-2456.

Zatorre RJ, Belin P. 2001. Spectral and temporal processing in human auditory cortex. Cereb.Cortex 11:946-953.

Zatorre RJ, Belin P, Penhune VB. 2002. Structure and function of auditory cortex: music and speech. Trends Cogn Sci. 6:37-46.

Zatorre RJ, Evans AC, Meyer E, Gjedde A. 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846-849.

4.4 Study G:

Electrical brain imaging evidences left auditory cortex involvement in speech and non-speech discrimination based on temporal features

TINO ZAEHLE [1], LUTZ JANCKE [1] & MARTIN MEYER [1,2]

[1]Departement of Neuropsychology, University of Zurich, Switzerland
[2]Institute of Neuroradiology, University Hospital of Zurich, Switzerland

Abstract

In the present study, we recorded and compared scalp auditory evoked potentials (AEP) in response to consonant-vowel-syllables with varying voice-onset-times (VOT) and non-speech analogues with varying noise-onset-time (NOT). In particular, we aimed to investigate the spatio-temporal pattern of acoustic feature processing underlying elemental speech perception and relate this temporal processing mechanism to specific activations of the auditory cortex.

Results show that the characteristic AEP waveform in response to consonant-vowel-syllables is on a par with those of non-speech sounds with analogue temporal characteristics. The amplitude of the N1a and N1b component of the auditory evoked potentials significantly correlated with the duration of the VOT in CV-syllables and likewise, with the duration of the NOT in non-speech sounds. Furthermore, current density maps indicate overlapping supratemporal networks involved in the perception of both speech and non-speech sounds with a bilateral activation pattern during the N1a time window and leftward asymmetry during the N1b time window. Elaborate regional statistical analysis of the activation over the middle and posterior portion of the supratemporal plane (STP) revealed strong left lateralized responses over the middle STP for both, the N1a and N1b component and a functional leftward asymmetry over the posterior STP for the N1b component.

The present data demonstrate overlapping spatio-temporal brain responses during the perception of temporal acoustic cues in both, speech and non-speech sounds. Source estimation evidences a preponderant role of the left middle and posterior auditory cortex in speech and non-speech discrimination based on temporal features. Therefore, in congruency with recent fMRI studies we suggest that similar mechanisms underlie the perception of linguistically different but acoustically equivalent auditory events on the level of basic auditory analysis.

Introduction

Auditory language perception is based on a variety of spectral and temporal acoustic information available in the speech signal (Davis and Johnsrude, 2007). One important temporal cue used to distinguish between stop-consonants is the voice onset time (VOT). The VOT, defined as the duration of the delay between release of closure and start of voicing, characterizes voicing differences among stop consonants in a wide variety of languages (Lisker and Abramson, 1964) and can thus be considered one of the most important acoustic cue encoding linguistically relevant information. The perceptual ability of resolving two signals as temporally discrete requires that the brain has a temporally segregated representation of those events.

Electrophysiological studies have consistently demonstrated VOT-related auditory evoked potential (AEP) differences in the N1 component with a single peak in response to short VOTs and with a double-peaked in response to longer VOTs for humans (Giraud *et al.,* 2005; Roman *et al.,* 2004; Sharma *et al.,* 2000; Sharma and Dorman, 1999; Steinschneider *et al.,* 1999), monkey (Steinschneider *et al.,* 1995; Steinschneider *et al.,* 2005), and guinea pig (McGee *et al.,* 1996). In humans it has been shown that non-speech sounds with related temporal characteristics as CV sounds resemble this pattern of acoustic temporal processing (Liegeois-Chauvel *et al.,* 1999). In particular using intracerebral depth electrodes, this study showed that the evoked responses of the left, but not the right primary auditory cortex are differential for the processing of voiced and voiceless consonants and their non-speech analogous.

Further support for a general mechanism for encoding and analysing successive temporal changes in acoustic signals has been evidenced by studies demonstrating that patients with acquired brain lesions and aphasia (Efron, 1963; Swisher and Hirsh, 1972), children with general language-learning disabilities (Tallal and Piercy, 1973; Tallal and Stark, 1981) and children and adults with dyslexia (Tallal, 1980) show impaired auditory processing of temporal information in nonverbal stimuli. Furthermore, children with reading disabilities are deficient in phoneme perception, which

is reflected by inconsistent labelling of tokens in VOT series (Tallal *et al.,* 1980; Tallal *et al.,* 1993), and these children also perform less consistently in labelling of tone onset time tokens (Breier *et al.,* 2001), and exhibit poorer auditory order thresholds (Von Steinbuchel, 1998). Moreover, it is known that the ability for phoneme discrimination in these children can be increased by a behavioural training using more salient versions of the rapidly changing elements in the acoustic waveform of speech (Merzenich *et al.,* 1996; Tallal *et al.,* 1996).

Recent electrophysiological and neuroimaging studies point to the important role of the primary and secondary auditory cortex for the processing of acoustic features in speech and nonspeech sounds. Several investigations using intracranial recording (Liegeois-Chauvel *et al.,* 1999; Steinschneider *et al.,* 2005), scalp EEG (Sandmann *et al.,* 2007; Trebuchon-Da *et al.,* 2005), MEG (Papanicolaou *et al.,* 2003) as well as fMRI (Meyer *et al.,* 2005; Zaehle *et al.,* 2004) demonstrated an elevated role of the human primary auditory cortex for the temporal processing of short acoustic cues in speech and non-speech sounds. Furthermore, also auditory association areas along the posterior supratemporal plane, in particular the bilateral planum temporale (PT) have been associated with the processing of rapidly changing auditory information during sub-lexical processing (Jancke *et al.,* 2002; Zaehle *et al.,* 2004). However, due to BOLD-related limitations in temporal resolutions, the EEG method is far more suitable to elucidate the temporal organization of speech perception. In combination with a recently developed source estimation algorithm (Pascual-Marqui *et al.,* 1999) it even allows to map the spatiotemporal dynamics of elemental aspects of speech perception, i.e. VOT decoding. Thus, the most important goal of this study is the validation of the aforementioned left middle and posterior auditory cortex recruitment in speech and non-speech discrimination based on temporal features.

In the present study, we recorded and compared scalp AEPs in response to CV-syllables and non-speech analogues with varying *VOT* and noise-onset-time (*NOT*), respectively. Here we aimed to investigate the neural

coding of acoustic characteristics underlying speech perception and relate this temporal processing mechanism to specific activations of the auditory cortex. It has been demonstrated that these processing mechanisms are reflected by modulations of the AEP. In particular, the N1 deflection is an obligatory component considered to reflect the basic encoding of acoustic information of the auditory cortex (Naatanen and Picton, 1987; Picton *et al.,* 1987). Furthermore this component reflects central auditory representation of speech sounds (Ostroff *et al.,* 1998; Sharma and Dorman, 2000) as well as of non-speech sounds (Pratt *et al.,* 2007). Thus, in the context of the present study we focused on the modulations during the N1 time window elicited by brief auditory stimuli that varied systematically along an acoustic and a linguistic dimension. In addition, we examined particularly to what extent the pattern of neural activation differs in distinct portions of the auditory cortex. As mentioned above both the middle compartment of the supratemporal plane (STP) accommodating the primary auditory cortex and the posterior compartment of the supratemporal plane harboring the planum temporale are crucial for processing transient acoustic features in speech and nonspeech sounds. In order to systematically investigate the contribution of these auditory cortex sections, we applied a low-resolution brain electromagnetic tomography (LORETA) approach and predicted functional leftward asymmetric responses over the middle and posterior portion of the STP to rapidly changing acoustic cues.

Material and Methods

Behavioural Pilot Study

In a behavioural pilot study, 24 healthy, right-handed native speakers of German (13 female, mean age = 26.7 ± 4.56 years) performed a phonetic categorization task. A synthetic VOT continuum was used ranging from 20 to 40 msec VOT in 1 msec steps. Participants were instructed to listen to each syllable and to decide whether the syllable was [da] or [ta], by pressing a corresponding button as quickly and accurately as possible. Figure 1

illustrates results of this pilot study. The graph shows the averaged identification curve indicating the percentage of VOT-continua syllables that were identified as /ta/. As indicated in Figure 1 the mean categorization boundary as indicated by the inflection point of the fitted polynomal function was at a VOT of 30 ms. Based on the results of these behavioural testing, for the electrophysiological investigation we used syllables with the VOT of 5 ms, that was consistently identified as the syllable /da/, with the VOT of 60 ms, that was consistently identified as the syllable /ta/, and with the VOT of 30 ms reflecting the averaged categorization boundary between /da/ and /ta/.

EEG Study

The electrophysiological experiment was conducted in a dimly lit, sound attenuated chamber. Subjects were placed in a comfortable chair at 110 cm distance from the monitor and scalp recorded event-related potentials (ERPs) in response to consonant – vowel (CV) syllables and nonspeech sounds were obtained from 18 male right handed, native German speaking healthy volunteers (mean age = 28.6 ± 3.45 years). None had any history of hearing, neurological, or psychiatric disorders. After a full explanation of the nature and risks of the study, subjects gave informed consent for the participation according to a protocol approved by the local Ethics Committee.

The auditory stimuli were generated with a sampling depth of 16 bits and a sampling rate of 44.1 kHz using the SoundForge 4.5 Software (Sonic Foundry Inc., http://www.sonicfoundry.com). We used a modified version of the stimulus material described by Zaehle *et al.*, (2004).

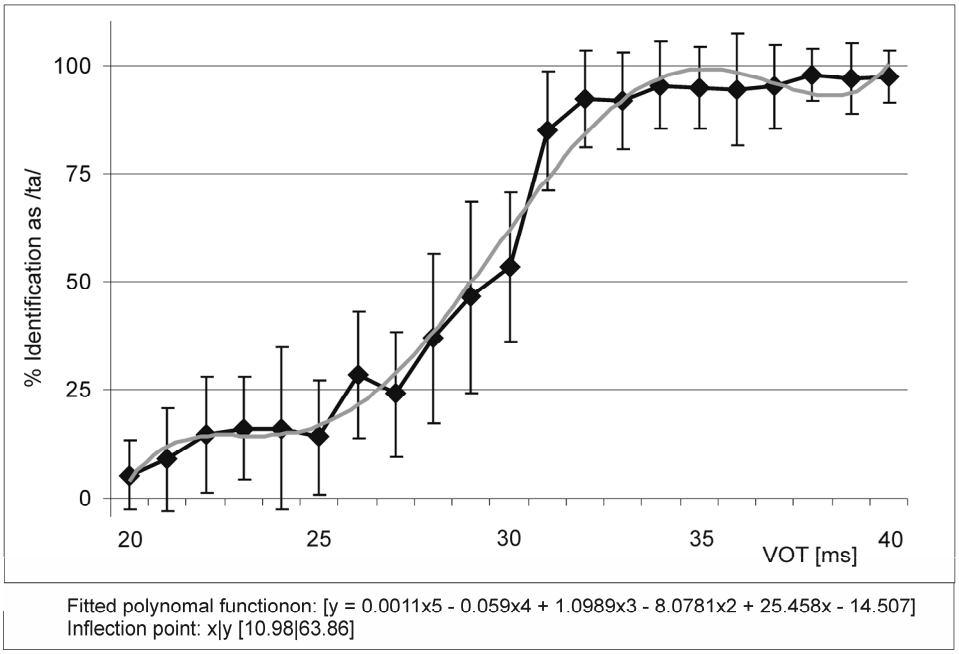

Figure 1: Averaged identification curve (+-1 standard deviation) indicating the percentage of CV-syllables that were identified as /ta/ in relation to their VOT (black, diamonds) and fitted polynomial function (gray)

Figure 2 shows wave-forms of the applied stimuli. Stimuli material consisted of CV syllables with varying voice-onset-times (5 ms, 30 ms and 60 ms) as revealed in the pilot behavioural study and analogously, non-speech sounds with varying noise-onset-times (5 ms, 30 ms and 60 ms). For the non-speech condition, we created stimuli containing two sound elements separated by a gap. The leading element was a wideband noise burst with a length of 7 ms. The trailing element was a bandpassed noise centred on 1.0 kHz and an width of 500 Hz. The duration of the gap was varied (5 ms, 30 ms, 60 ms). The duration of each single stimulus was consistent (330 ms). Auditory stimuli were presented binaurally using Hifi-headphones (55 dB sound pressure level). Stimulation and recording of responses were controlled by the Presentation software (Neurobehavioral Systems, USA).

In the EEG experiment, each session comprised ten blocks. Within each block, 18 trials of each stimulus category were presented in a randomized order resulting in a total number of 180 stimulus presentations

during the entire experimental session. For each trial volunteers performed a same-different discrimination task on pairs of stimuli belonging to one stimulus category. Stimuli varied with respect to the temporal manipulation of the *NOT* and *VOT*. Stimuli of one pair were presented with an inter stimulus interval of 1300 ms. Participants indicated their answers by pressing one of two response buttons. We utilized this task to ensure subjects' vigilance throughout the experiment and to engage the subjects to attend to the auditory stimulation. However, we were primarily interested in the electrophysiological responses to acoustic features underlying pure and elemental speech perception. Respectively, we aimed to avoid confound with neural correlates of decision making instantly following the second stimulus of each pair of VOT and NOT. Thus, only the first stimulus of each stimulus pair was analysed and included into the following analysis.

Electroencephalogram (EEG) was recorded from 32 scalp electrodes (30 channels + 2 eye channels) located at standard left and right hemisphere positions over frontal, central, parietal, occipital, and temporal areas (subset of international 10/ 10 system sites: Fz, FCz, Cz, CPz, Pz, Oz, Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, TP7,TP8, FT7, FT8, FC3, FC4, CP3, and CP4) using a band pass of 0.53 -70 Hz with a sampling rate of 500 Hz. We applied sintered silver/silver chloride electrodes (Ag/AgCl) and used the FCz position as reference. Impedances of these electrodes were kept below 5 kΩ. Trials containing ocular artifacts, movement artifacts, or amplifier saturation were excluded from the averaged ERP waveforms. The processed data were re-referenced to a virtual reference derived from the average of all electrodes. Each ERP waveform was an average of more than 100 repetitions of the potentials evoked by the same stimulus type. The EEG recordings were sectioned into 600 ms epochs (100 ms pre-stimulus and 500 ms post-stimulus) and a baseline correction using the prestimulus portion of the signal was carried out. ERPs for each stimulus were averaged for each subject and grand-averaged across subjects.

In order to statistically confirm the predicted differences between AEP components at Cz as a function of experimental stimuli, mean amplitude ERPs time-locked to the auditory stimulation were measured in two latency

windows (110-119 msec and 190-209 msec) determined by both visual inspection and based on previous results covering the prominent N1 and N2 components. Analyses of variance (ANOVAs) with factors *temporal modulation* (5, 30, 60 ms) and *speechness* (VOT/NOT) were computed for central electrode (Cz), and the p values reported were adjusted with the Greenhouse–Geisser epsilon correction for nonsphericity.

Subsequently, we applied an inverse linear solution approach – LORETA (low-resolution electromagnetic tomography) to estimate the neural sources of event-related scalp potentials (Pascual-Marqui *et al.,* 1994; Pascual-Marqui *et al.,* 2002). In order to verify the estimated localization of the N1 and N2 component, we calculated the LORETA current density value ($\mu A/mm^2$) for the AEPs within the 3D voxel space. We used a transformation matrix with high regularization (1e3 * (first eigenvalue)) to increase signal to noise ratio. The maxima of the current density distributions were displayed on a cortical surface model and transformed in stereotactic Talairach space (Talairach and Tournoux, 1988). Subsequently, to specifically test neurofunctional hypothesis of the bilateral middle and posterior STP, we calculated a post hoc region-of-interest (ROI) analysis. We defined four 3D ROIs in STP (left middle STP, right middle STP, left posterior STP, right posterior STP). The landmarks of ROIs were determined by an automatic anatomical labeling procedure implemented in LORETA. By means of the ROI extractor software tool[4], we collected mean current density values from each individual and each distinct 3D ROI. The mean current density values for each ROI were submitted to a 3x2x2 ANOVA with the factors *temporal modulation* (5, 30, 60 ms), *hemisphere* (left/right) and *speechness* (VOT/NOT).

[4] (http://www.novatecheeg.com/Downloads.html)

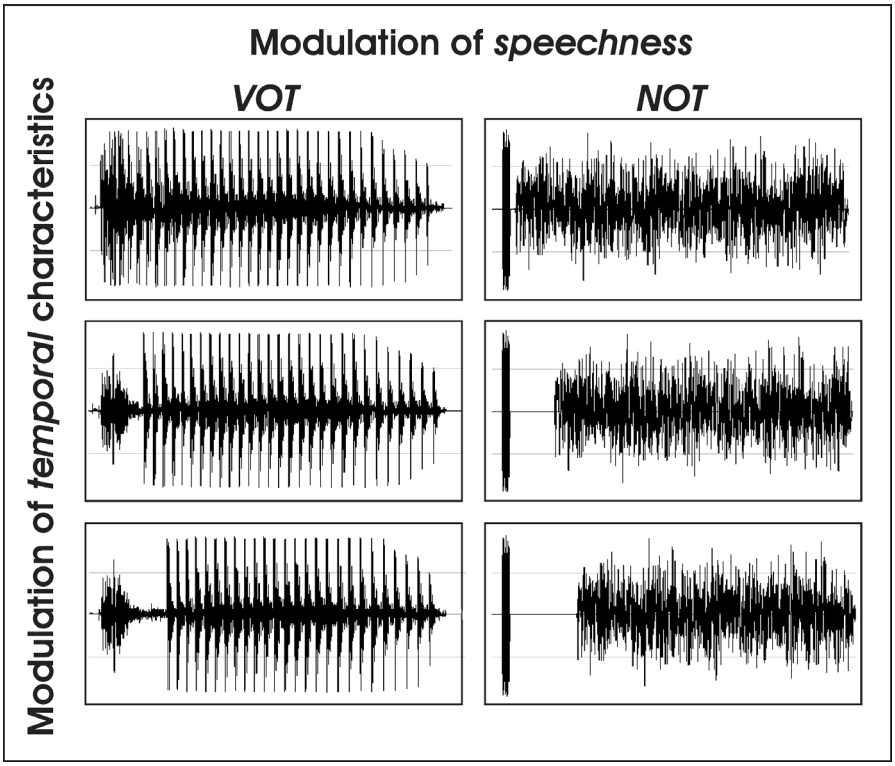

Figure 2: Waveforms of auditory stimuli. The left panel shows speech stimuli (CV) with varying VOT (5, 30, 60 ms), and the right panel shows non-speech stimuli with varying NOT (top to bottom: 5, 30, 60 ms).

Results

Electrophysiological Data

Grand averaged waveforms evoked by each of the four speech and non-speech stimuli recorded from Cz are shown in Figure 3. As one can see from this figure we observed that all stimuli elicited a prominent N1a component with the shortest VOT/NOT modulation (5 ms) yielding the most enhanced amplitude. Furthermore we noticed a second negative deflection peaking around 200 ms after stimulus onset (N1b) also revealing sensitivity to the temporal modulation of the sounds. In order to statistically examine the ERP effects, mean amplitude of the ERP waveforms were measured in two 20-msec latency windows.

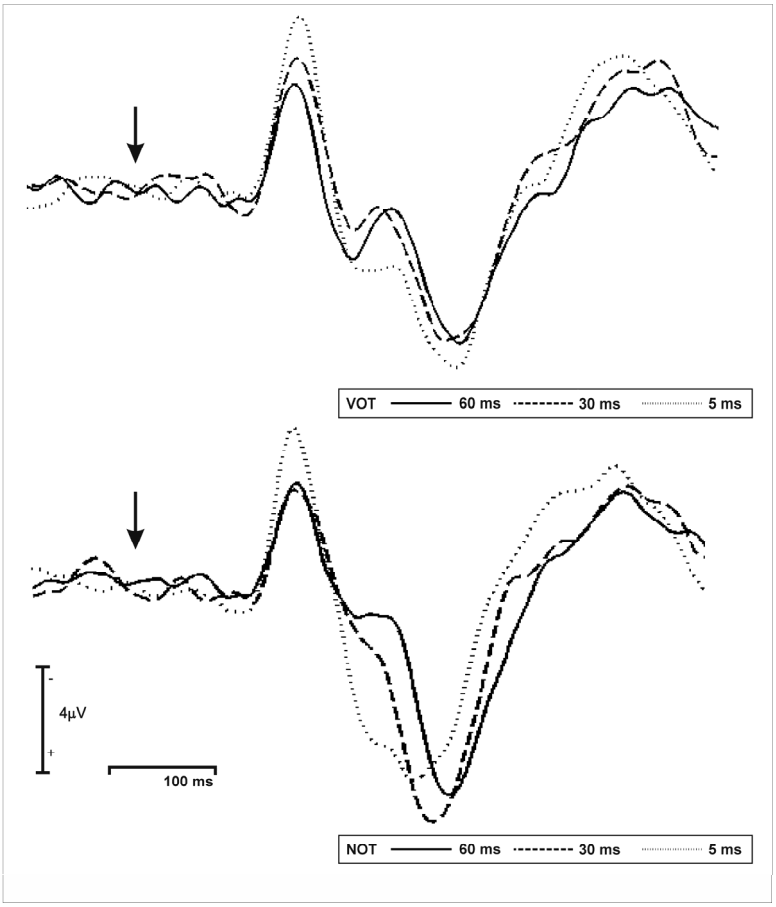

Figure 03: Averaged electrophysiological data, recorded from 18 participants time locked at the onset of stimulation at central (Cz) electrode during the perception of VOT (top) and NOT stimuli.

Results of the 3x2 ANOVA with the factors *temporal modulation* (5, 30, 60 ms) and *speechness* (VOT/NOT) for the N1a (TW I: 110-129 ms latency window) revealed a significant main effect of the factor *temporal modulation* ($F(1.77, 30.1) = 12.45$, $p < 0.001$). Similar, for the N1b (190-209 ms latency window) ANOVA revealed a significant main effect of the factor *temporal modulation* ($F(1.58, 26.92) = 15.7$, $p < 0.001$). Furthermore, ANOVA for the N1b also revealed a significant main effect of the factor *speechness* ($F(1, 17) = 19.88$, $p < 0.001$), and a significant *temporal modulation* by *speechness* interaction ($F(1.6, 27.4) = 4.79$, $p < 0.05$).

Subsequently, *post-hoc* analyses were conducted separately for speech and non-speech stimulation. Figure 4 shows plots of mean amplitude of the temporal modulation separated for speech and nonspeech for a) N1a and b) N1b. Results of the one-factorial ANOVAs are listed in Table 1. For the N1 (110-129 ms latency), separate one-factorial ANOVA revealed a significant

main effect of the factor *temporal modulation* for the non-speech sounds ($F(1.8, 30.9) = 8.14$ $p < 0.001$). Test for linear contrast demonstrate significant linear relationship of the N1a mean amplitude and length of the *NOT* in the non-speech sounds ($F(1,17) = 15.53$, $p=0.001$). Similarly, one-factorial ANOVA with the factor *temporal modulation* in the speech sounds revealed a significant main effect ($F(1.61, 27.4) = 5.34$, $p < 0.05$), and test for linear contrast revealed significant linear relationship of the N1a mean amplitude and length of the *VOT* in the speech sounds ($F(1,17) = 9.39$, $p < 0.05$). The same pattern of activation was present at the 190 – 209 ms latency window (N1b). Separate one- factorial ANOVAs revealed a significant main effect of the factor *temporal modulation* for the non-speech sounds ($F(1.23, 21.1) = 18.09$, $p < 0.001$), and one-factorial ANOVA with the factor *temporal modulation* revealed a significant main ($F(1.79, 30.49) = 3.85$, $p < 0.05$) for the speech sounds. Tests for linear contrast revealed significant linear relationship of the N1b mean amplitude and length of the *NOT* in the non-speech sounds ($F(1,17) = 24.18$, $p < 0.001$), and *VOT* in the speech sounds ($F(1,17) = 4.99$, $p < 0.05$).

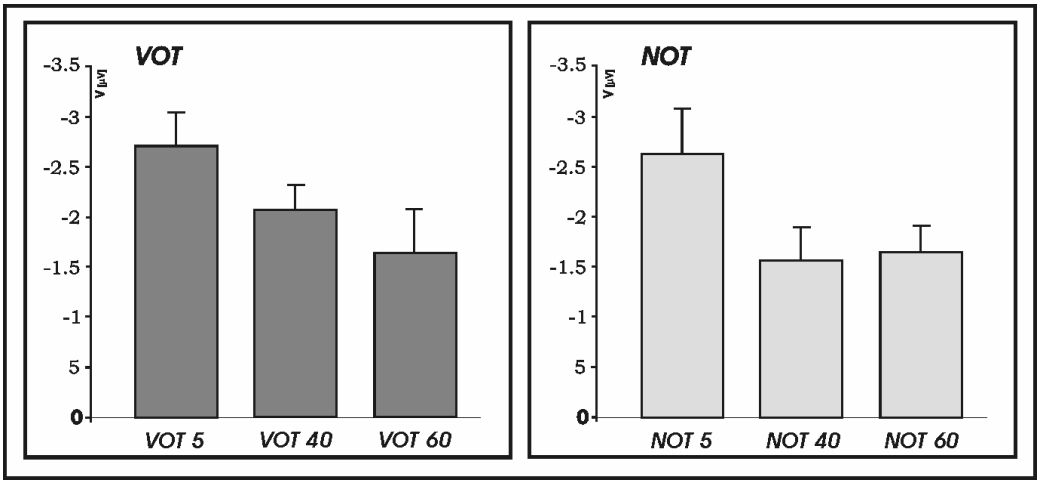

Figure 04a: Plots of mean amplitude for N1a separate for *VOT* and *NOT* stimuli

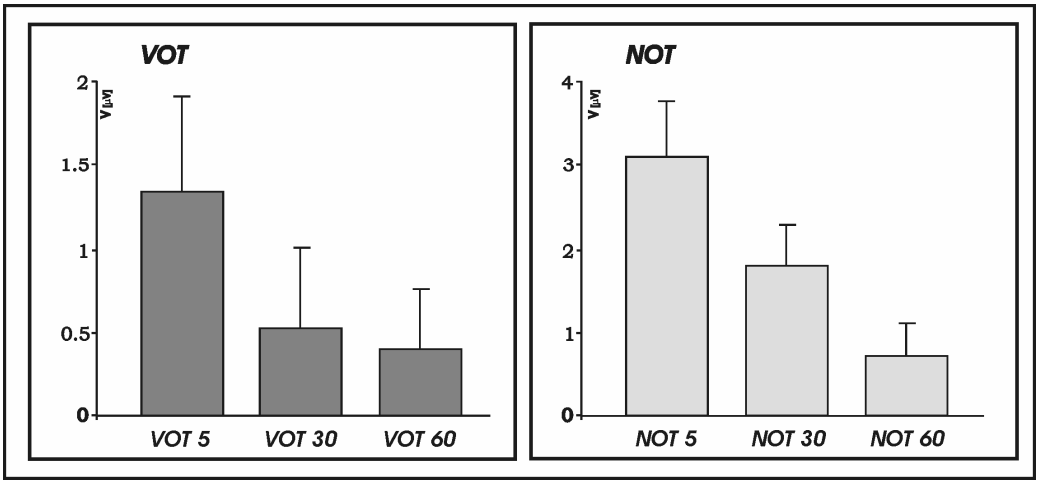

Figure 04b: Plots of mean amplitude for N1b separate for *VOT* and *NOT* stimuli

Source localisation

Table 2 presents coordinates and corresponding brain regions associated with current density maxima for the speech and non-speech sounds obtained separately for the N1a and N1b time windows. As shown in Figure 5, for the N1a time window current density maps indicate that left and right posterior perisylvian areas contribute to both, speech and non-speech sounds.

Table 1: Results of ANOVAs with the factor NOT and VOT for TW I and TW II

Factor				**linear contrast**		
	df	**F-value**	**p - value**	**df**	**F-value**	**p - value**
Time window I (N1)						
VOT	1.61	5.34	0.01	1	9.39	0.007
NOT	1.81	8.14	0.001	1	15.53	0.001
Time window II (N2)						
VOT	1.79	3.84	0.03	1	4.98	0.04
NOT	1.24	18.09	0.000	1	24.18	0.000

With regard to the N1b, source estimation showed enlarged current density distribution over the left posterior STP and the anterior cingulate

gyrus for speech and non-speech sounds, and the right posterior STP for non-speech sounds.

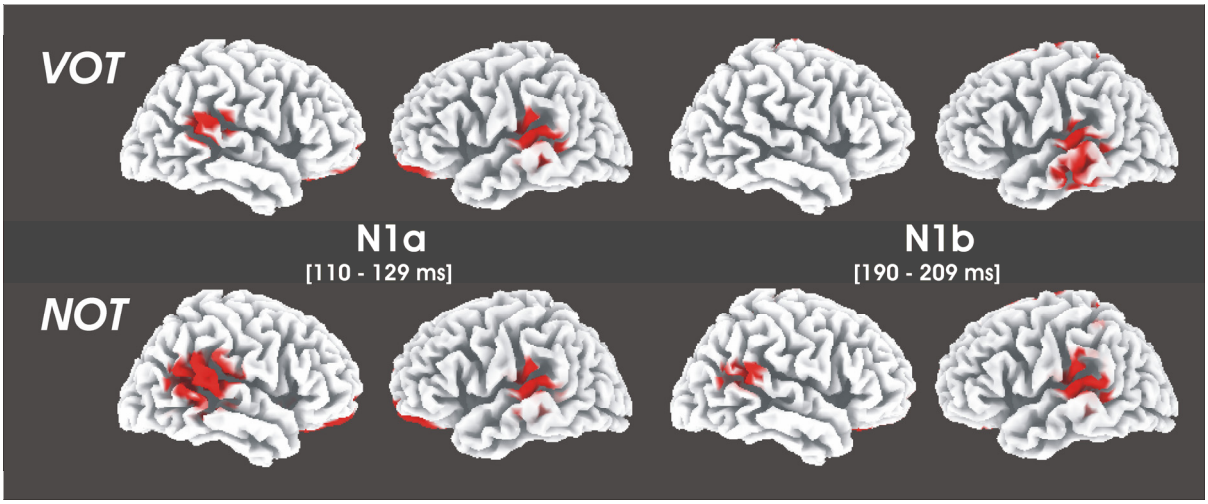

Figure 05: Grand average (n=18) three dimensional LORETTA –based current density maxima for AEP components N1 and N2. (Threshold: 0.001 prop. µA/mm2)

Subsequent statistical analysis of ROIs over the bilateral middle portion of the STP separate for N1a and N1b time windows revealed that current density values were strongly lateralized. A 3x2x2 ANOVA with the factors *temporal modulation* (5, 30, 60 ms), *hemisphere* (left/right) and *speechness* (VOT/NOT) revealed a significant main effect of the factor *hemisphere* ($F(1,17)=18.64$, $p < 0.001$) for the N1a as well as for the N1b time window ($F(1,17)=27.97$, $p < 0.001$), demonstrating stronger responses over the left as compared to the right primary auditory cortex. Figure 6 shows current density values during the processing of VOT and NOT stimuli collapsed over the temporal modulations and extracted from the left and right primary auditory cortex.

Table 2: Current density maxima [µA/mm^2]*10^{-3} in response to speech (VOT) and non-speech (NOT) sounds

Component	Condition	Brain Region	Current density value	Hemisphere	X	Y	Z
N1a	VOT	Cingulum	1.74		-3	45	1
		STG	1.39	L	-59	-32	8
			1.30	R	60	-39	15
	NOT	Cingulum	2.70		-3	45	1
		STG	1.50	L	-59	-32	8
			1.78	R	60	-39	15
N1b	VOT	Cingulum	1.74		-3	45	1
		STG	1.39	L	-59	-32	8
	NOT	Cingulum	2.70		-3	52	1
		STG	1.50	L	-59	-32	8
			1.78	R	60	-39	15

Analysis for the posterior portion of the STP showed no significant main effect or an interaction for the N1a time window. For the N1b time window, analysis showed a significant main effect of the factor *hemisphere* ($F(1,17)=5.55$, $p < 0.05$) indicating stronger responses over the left as compared to the right posterior STP.

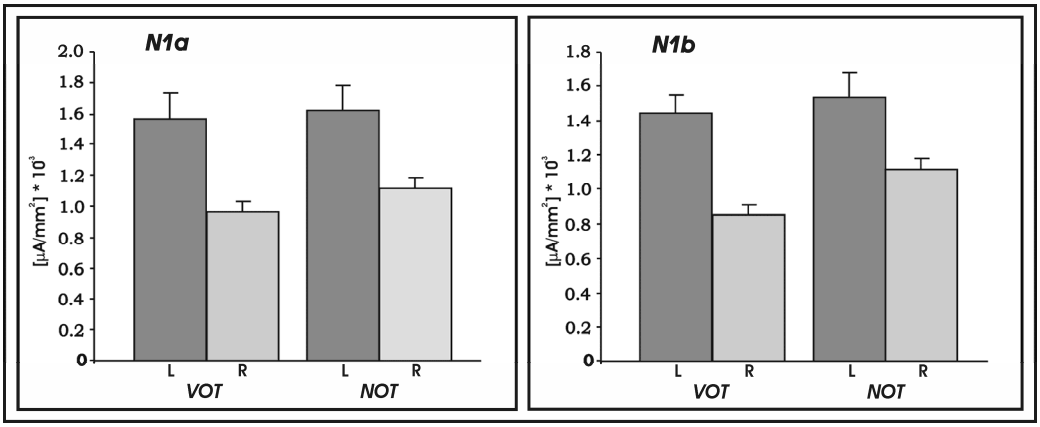

Figure 06: Plots of mean current density values obtained by the anatomically defined ROI analysis, separate for the left and right middle portion of the supratemporal plane (BA41): Left panel shows date for N1a (TW I) and the right panel shows data for N1b (TW II).

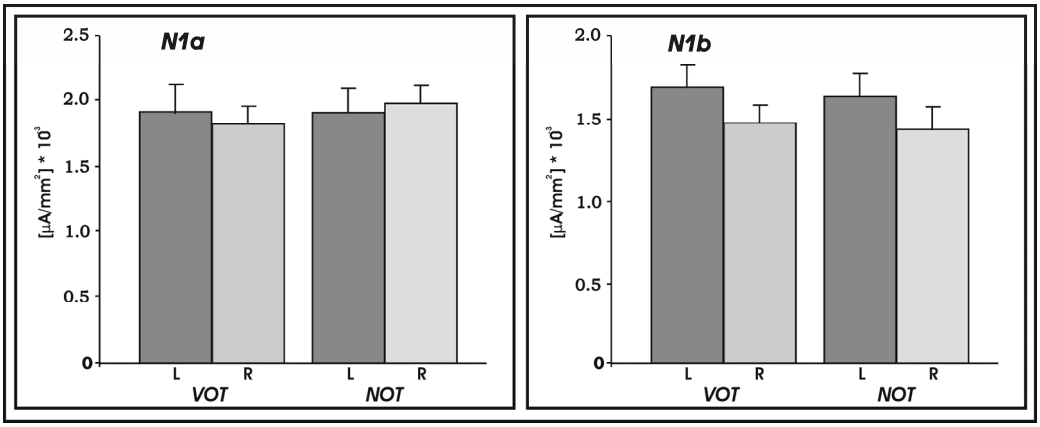

Figure 07: Plots of mean current density values obtained by the anatomically defined ROI analysis, separate for the left and right posterior portion of the supratemporal plane (post BA42): Left panel shows date for N1a (TW I) and the right panel shows data for N1b (TW II).

Discussion

One of the key questions in understanding the nature of speech perception is to what extent the human brain has unique speech-specific mechanisms or to what degree it processes sounds equally depending on their acoustic properties. In the present study we showed that the characteristic AEP waveform in response to consonant-vowel-syllables shows an almost identical spatio-temporal pattern as in response to non-speech sounds with similar temporal characteristics. The amplitudes of the N1a and N1b component of the auditory evoked potentials significantly correlated with the duration of the VOT in CV-syllables and analogously, with the duration of the NOT in non-speech sounds. Furthermore, current density maps of the N1a and N1b time windows indicate overlapping neural distribution of these components originating from the same sections over the superior temporal plane that accommodates auditory cortex. For the analysis of the middle portion of the STP incorporating the primary auditory cortex, we revealed asymmetric activations that point to a stronger involvement of left supratemporal plane regardless of TW, speechness or temporal modulation. For the posterior part of the STP, the analysis of the current density values revealed a bilateral activation pattern during the N1a

time window and a leftward asymmetry during the N1b time window for both, the perception of speech and non-speech sounds.

In general or data are in line with former electrophysiological studies investigating processing of brief auditory cues but delivers novel insight in that it demonstrates a strong preference of the left middle and posterior auditory cortex for rapidly modulating temporal information by means of a low-resolution source estimation approach. Using MEG it has been demonstrated that AEP response to speech sounds exhibit a N100m which is followed by a N200m at around 200-210 ms (Kaukoranta et al., 1987). It has been proposed that the N200m is specific to acoustic parameters available in vowels, since the N200m was triggered by acoustic rather than phonetic features of the stimulus. Sharma and colleges showed that the typical change in the AEP waveform morphology from single to double peaked N1 component is not a reliable indicator of perception of voicing contrasts in syllable-initial position (Sharma et al., 2000). In other words, a double–peak onset response cannot be considered a cortical correlate of the perception of voicelessness, it rather depends on the acoustic properties of the sound signal. For the perception of consonants with the same place of articulation, the critical acoustic feature which distinguishes between these consonants is the time between the burst at consonant initiation and the onset of voicing (VOT). Similarly, in case of the non-speech sounds the critical acoustic feature is the time (silent gap) between the trailing and leading noise elements. In both cases, the ability to perform the task requires the listener to perceptually segregate the two sounds (or their onsets) in time, which in turn requires that the brain have temporally segregated responses to the two events (or their onsets) (Phillips, 1993). As demonstrated by the present data, overlapping cortical excitement was found for the detection of temporal cues in both speech and non-speech sounds. Therefore, our data support the notion of similar mechanisms underling the perception auditory events that are equal in temporal acoustic structure but differ in linguistic meaning.

It has been suggested that the primary auditory cortex is specifically involved in the perceptual elaboration of sounds with durations or spacing

within a specific temporal grain (Phillips, 1993) and this suggestion has been confirmed by studies demonstrating that primary auditory cortex evoked responses reflect encoding of VOT (Liegeois-Chauvel et al., 1999; Sandmann et al., 2007; Steinschneider et al., 2005; Trebuchon-Da et al., 2005). Furthermore, the Heschl's gyrus (HG) is known to display a leftward structural asymmetry (Dorsaint-Pierre et al., 2006; Penhune et al., 1996; Penhune et al., 2003; Rademacher et al., 1993). This asymmetry is related to a larger white matter volume of the left as compared to the right HG (Penhune et al., 1996; Sigalovsky et al., 2006) as well as to asymmetries at the cellular level (Hutsler and Gazzaniga, 1996; Seldon, 1981a; Seldon, 1981b; Seldon, 1982). It has been hypothesised that this leftward asymmetry of the HG is related to a more efficient processing of rapidly changing acoustic information, which is relevant in speech perception (Zatorre and Belin, 2001).

The posterior part of the left STP that partly covers the planum temporale (PT) has also been associated with competence to mediate spectro-temporal integration during auditory perception (Griffiths and Warren, 2002; Warren et al., 2005). In particular the left posterior auditory cortex plays a prominent role when speech relevant auditory information has to be processed (Geiser et al., 2007; Meyer et al., 2005; Zaehle et al., 2004). Akin to the primary auditory cortex that resides in HG, the posterior STP also has structural leftward asymmetry (Anderson et al., 1999; Galuske et al., 2000), which indicates a relationship between this brain region and the leftward lateralized, specific functions relevant for speech perception.

As mentioned above, a long lasting question in auditory speech research concerns the nature of the VOT cue and asks to what extent the VOT is processed by specialized speech mechanisms or by more basic acoustically tuned mechanisms (Pisoni, 1977). As suggested by Phillips (1993), as long as the stimulus representation in the primary auditory cortex is concerned, speech may be "special" only in the sense that spoken language is the most obvious stimulus in which the identification of the elements is dependent on temporal resolution (Phillips, 1993). In fact, data of the present study evidence that the middle and posterior auditory cortex

especially of the left hemisphere is significantly involved in the processing of the acoustically features critical for the processing of temporal cues in both, speech and non-speech sounds.

This conclusion corroborates recent fMRI research, but in addition demonstrates that EEG in combination with low-resolution tomography could be considered an ideal alternative to map the spatio-temporal patterns of speech perception. In a way, this approach outperforms the fMRI technology because it evidently demonstrates the temporal subtlety of elemental acoustic processing reflected by differential sensitivity and neural distribution of succeeding N1a and N1b responses to brief speech and speech-like stimuli. Of course, one should bear in mind that spatial resolution of electrophysiologically based localization methods is inferior to modern brain imaging techniques. Thus, by no means one should feel tempted to interpret the activation maps provided by LORETA in an fMRI-like manner. However, it has been proven that low-resolution tomography is capable of reliably distinguishing between sources originating from distinct sections of the superior temporal region (Meyer et al., 2006). This hold particularly true if low-resolution tomography is used to examine electrophysiological responses emerging from the left or right hemisphere (Sinai and Pratt, 2003).

In essence, the present study delivers further evidence for the prominent role of the middle and posterior left supratemporal plane in the perception of rapidly changing cues which is thought to be an essential device underlying speech perception (Hickok and Poeppel, 2007; Poeppel, 2003; Zatorre and Belin, 2001).

Reference List

Anderson, B., Southern, B.D., Powers, R.E., 1999. Anatomic asymmetries of the posterior superior temporal lobes: a postmortem study. Neuropsychiatry Neuropsychol.Behav.Neurol. 12, 247-254.

Breier, J.I., Gray, L., Fletcher, J.M., Diehl, R.L., Klaas, P., Foorman, B.R., Molis, M.R., 2001. Perception of voice and tone onset time continua in children with dyslexia with and without attention deficit/hyperactivity disorder. J.Exp.Child Psychol. 80, 245-270.

Davis, M.H., Johnsrude, I.S., 2007. Hearing speech sounds: Top-down influences on the interface between audition and speech perception. Hear.Res. 229, 132-147.

Dorsaint-Pierre, R., Penhune, V.B., Watkins, K.E., Neelin, P., Lerch, J.P., Bouffard, M., Zatorre, R.J., 2006. Asymmetries of the planum temporale and Heschl's gyrus: relationship to language lateralization. Brain 129, 1164-1176.

Efron, R., 1963. Temporal Perception, Aphasia and D'ej'a vu. Brain 86, 403-424.

Galuske, R.A., Schlote, W., Bratzke, H., Singer, W., 2000. Interhemispheric asymmetries of the modular structure in human temporal cortex. Science 289, 1946-1949.

Geiser, E., Zaehle, T., Jancke, L., Meyer, M., 2007. The neural correlate of speech rhythm as evidenced by meter processing: an fmri study.

Giraud, K., Demonet, J.F., Habib, M., Marquis, P., Chauvel, P., Liegeois-Chauvel, C., 2005. Auditory evoked potential patterns to voiced and voiceless speech sounds in adult developmental dyslexics with persistent deficits. Cereb.Cortex 15, 1524-1534.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci. 25, 348-353.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat.Rev.Neurosci. 8, 393-402.

Hutsler, J.J., Gazzaniga, M.S., 1996. Acetylcholinesterase staining in human auditory and language cortices: regional variation of structural features. Cereb.Cortex 6, 260-270.

Jancke, L., Wustenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. Neuroimage. 15, 733-746.

Kaukoranta, E., Hari, R., Lounasmaa, O.V., 1987. Responses of the human auditory cortex to vowel onset after fricative consonants. Exp.Brain Res. 69, 19-23.

Liegeois-Chauvel, C., de Graaf, J.B., Laguitton, V., Chauvel, P., 1999. Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb.Cortex 9, 484-496.

Lisker, L., Abramson, A.S., 1964. Across language study of voicing in initial stops: Acoustical measurements. Word 20, 384-411.

McGee, T., Kraus, N., King, C., Nicol, T., Carrell, T.D., 1996. Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. J.Acoust.Soc.Am. 99, 3606-3614.

Merzenich, M.M., Jenkins, W.M., Johnston, P., Schreiner, C., Miller, S.L., Tallal, P., 1996. Temporal processing deficits of language-learning impaired children ameliorated by training. Science 271, 77-81.

Meyer, M., Baumann, S., Jancke, L., 2006. Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. Neuroimage 32, 1510-1523.

Meyer, M., Zaehle, T., Gountouna, V.E., Barron, A., Jancke, L., Turk, A., 2005. Spectro-temporal processing during speech perception involves left posterior auditory cortex. Neuroreport 16, 1985-1989.

Naatanen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24, 375-425.

Ostroff, J.M., Martin, B.A., Boothroyd, A., 1998. Cortical evoked response to acoustic change within a syllable. Ear Hear. 19, 290-297.

Papanicolaou, A.C., Castillo, E., Breier, J.I., Davis, R.N., Simos, P.G., Diehl, R.L., 2003. Differential brain activation patterns during perception of voice and tone onset time series: a MEG study. Neuroimage. 18, 448-459.

Pascual-Marqui, R.D., Esslen, M., Kochi, K., Lehmann, D., 2002. Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. Methods Find.Exp.Clin.Pharmacol. 24 Suppl C, 91-95.

Pascual-Marqui, R.D., Lehmann, D., Koenig, T., Kochi, K., Merlo, M.C., Hell, D., Koukkou, M., 1999. Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. Psychiatry Res. 90, 169-179.

Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int.J.Psychophysiol. 18, 49-65.

Penhune, V.B., Cismaru, R., Dorsaint-Pierre, R., Petitto, L.A., Zatorre, R.J., 2003. The morphometry of auditory cortex in the congenitally deaf measured using MRI. Neuroimage 20, 1215-1225.

Penhune, V.B., Zatorre, R.J., MacDonald, J.D., Evans, A.C., 1996. Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. Cereb.Cortex 6, 661-672.

Phillips, D.P., 1993. Neural representation of stimulus times in the primary auditory cortex. Ann.N.Y.Acad.Sci. 682, 104-118.

Picton, T.W., Skinner, C.R., Champagne, S.C., Kellett, A.J., Maiste, A.C., 1987. Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. J.Acoust.Soc.Am. 82, 165-178.

Pisoni, D.B., 1977. Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. J.Acoust.Soc.Am. 61, 1352-1361.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication 41, 245-255.

Pratt, H., Starr, A., Michalewski, H.J., Bleich, N., Mittelman, N., 2007. The N1 complex to gaps in noise: effects of preceding noise duration and intensity. Clin.Neurophysiol. 118, 1078-1087.

Rademacher, J., Caviness, V.S., Jr., Steinmetz, H., Galaburda, A.M., 1993. Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. Cereb.Cortex 3, 313-329.

Roman, S., Canevet, G., Lorenzi, C., Triglia, J.M., Liegeois-Chauvel, C., 2004. Voice onset time encoding in patients with left and right cochlear implants. Neuroreport 15, 601-605.

Sandmann, P., Eichele, T., Specht, K., Jancke, L., Rimoal, L.M., Nordby, K., Hugdahl, K., 2007. Hemispheric asymmetries in processing of temporal cues in consonant-vowel syllables.

Seldon, H.L., 1981a. Structure of human auditory cortex. I. Cytoarchitectonics and dendritic distributions. Brain Res. 229, 277-294.

Seldon, H.L., 1981b. Structure of human auditory cortex. II. Axon distributions and morphological correlates of speech perception. Brain Res. 229, 295-310.

Seldon, H.L., 1982. Structure of human auditory cortex. III. Statistical analysis of dendritic trees. Brain Res. 249, 211-221.

Sharma, A., Dorman, M.F., 1999. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. J.Acoust.Soc.Am. 106, 1078-1083.

Sharma, A., Dorman, M.F., 2000. Neurophysiologic correlates of cross-language phonetic perception. J.Acoust.Soc.Am. 107, 2697-2703.

Sharma, A., Marsh, C.M., Dorman, M.F., 2000. Relationship between N1 evoked potential morphology and the perception of voicing. J.Acoust.Soc.Am. 108, 3030-3035.

Sigalovsky, I.S., Fischl, B., Melcher, J.R., 2006. Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: a possible marker for primary cortex and hemispheric differences. Neuroimage 32, 1524-1537.

Sinai, A., Pratt, H., 2003. High-resolution time course of hemispheric dominance revealed by low-resolution electromagnetic tomography. Clin.Neurophysiol. 114, 1181-1188.

Steinschneider, M., Reser, D., Schroeder, C.E., Arezzo, J.C., 1995. Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. Brain Res. 674, 147-152.

Steinschneider, M., Volkov, I.O., Fishman, Y.I., Oya, H., Arezzo, J.C., Howard, M.A., III, 2005. Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. Cereb.Cortex 15, 170-186.

Steinschneider, M., Volkov, I.O., Noh, M.D., Garell, P.C., Howard, M.A., III, 1999. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. J.Neurophysiol. 82, 2346-2357.

Swisher, L., Hirsh, I.J., 1972. Brain damage and the ordering of two temporally successive stimuli. Neuropsychologia 10, 137-152.

Tallal, P., 1980. Auditory temporal perception, phonics, and reading disabilities in children. Brain Lang 9, 182-198.

Tallal, P., Miller, S., Fitch, R.H., 1993. Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann.N.Y.Acad.Sci. 682, 27-47.

Tallal, P., Miller, S.L., Bedi, G., Byma, G., Wang, X., Nagarajan, S.S., Schreiner, C., Jenkins, W.M., Merzenich, M.M., 1996. Language comprehension in language-learning impaired children improved with acoustically modified speech. Science 271, 81-84.

Tallal, P., Piercy, M., 1973. Defects of non-verbal auditory perception in children with developmental aphasia. Nature 241, 468-469.

Tallal, P., Stark, R.E., 1981. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. J.Acoust.Soc.Am. 69, 568-574.

Tallal, P., Stark, R.E., Kallman, C., Mellits, D., 1980. Developmental dysphasia: relation between acoustic processing deficits and verbal processing. Neuropsychologia 18, 273-284.

Trebuchon-Da, F.A., Giraud, K., Badier, J.M., Chauvel, P., Liegeois-Chauvel, C., 2005. Hemispheric lateralization of voice onset time (VOT) comparison between depth and scalp EEG recordings. Neuroimage 27, 1-14.

Von Steinbuchel, N., 1998. Temporal ranges of central nervous processing: clinical evidence. Exp.Brain Res. 123, 220-233.

Warren, J.D., Jennings, A.R., Griffiths, T.D., 2005. Analysis of the spectral envelope of sounds by the human brain. Neuroimage 24, 1052-1057.

Zaehle, T., Wustenberg, T., Meyer, M., Jancke, L., 2004. Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur.J.Neurosci. 20, 2447-2456.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb.Cortex 11, 946-953.

5. Concluding discussion

5.1 Methodological investigations

The first main goal of the present work was the implementation and evaluation of "clustered-sparse temporal acquisitions (CTA)" protocols in auditory fMRI research. For this purpose we performed two fMRI experiments to compare the new CTA protocol to a conventional continuous fMRI scanning scheme and a single-scan "sparse-temporal acquisition (STA)" scheme.

In the first study, entitled "*Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task*" we investigated the effects of ambient scanner noise bound to continuous fMRI on hemodynamic responses to spoken sentences by comparing conventional noisy fMRI with a silent clustered acquisition scheme. In the results we observed that activation in perisylvian regions corresponding to speech perception varied as a function of either the "noisy" or the "silent" scanning approaches. While auditory sentence processing in a silent fMRI paradigm brought on stronger functional responses of the bilateral supratemporal plane accommodating the entire auditory cortex, the presentation of the identical auditory stimulus material in the context of a traditional continuous acquisition recruited the bilateral insulae more strongly. This investigation clearly demonstrated that "clustered-sparse temporal acquisitions" protocols can successfully be applied to map functional responses to auditory sentence stimuli in a silent environment. At the perceptual level we revealed faster response times during "silent fMRI", which can be taken as evidence for facilitated auditory sentence processing. At the physiological level we show that "silent fMRI" is generally associated with stronger functional activation in the auditory cortices, as this approach is not limited by saturation of the neuronal population evoked by persistent

scanner noise. Additionally, we noticed noise dependent modulations of functional hemisphere asymmetry in the mid and posterior STP.

In the second study, entitled "*Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks*", we evaluated a clustered (CTA) and a sparse temporal acquisition protocol (STA). Here two event-related auditory experiments were performed during which the subjects were presented with short tonal and auditory sentence stimuli. The comparison of the two fMRI acquisition schemes revealed a clear pattern of results. At the level of single subject analyses, the CTA scheme clearly outperformed the STS scheme. In both the left and right temporal cortex the auditory stimulation related number of suprathreshold voxels was larger, and the maximum t-value was significantly higher for the CTA- than the STA scheme. This effect was true for relative long speech and relative short tonal stimulation. In the context of a statistical group analysis, results differentially benefited from the application of either a CTA- or STA-protocol. Whereas in the context of a random effects approach the comparison revealed only a marginal difference between the two acquisition schemes, fixed effects analyses gained a tremendous benefit by the use of a CTA scheme. These differences originate from contrary asymmetries in two related statistical parameters. While we observed an asymmetry in the amount of beta-values in favor of the STA scheme, an asymmetry in the height of t-values in favor of the CTA scheme became evident. To summarize, this study shows that regardless of the stimulus duration it is advantageous to acquire several sample points per trial if one wants to use the benefit of "silent" fMRI. Furthermore, data demonstrate that the acquisition of a cluster of subsequent imaging volumes along the T1-decay can be successfully applied in the context of auditory fMRI research.

5.2 Investigation

The second main goal of the present work was to examine the functional neuroanatomy of sub-lexical auditory perception with specific respect to auditory spectro-temporal processing in speech and non-speech sounds. For this purpose we conducted a series of four experiments (three fMRI and one EEG study) to investigating the nature of temporal and spectral processing in the auditory domain. The first fMRI experiment entitled "*Spectral-temporal processing during speech perception involves left posterior auditory cortex*" investigated the neural underpinnings of spectral-temporal integration during speech perception. The study demonstrated that the left posterior superior temporal lobe is essential for spectral-temporal processing during speech perception. The second fMRI experiment entitled "*The neural correlate of speech rhythm as evidenced by meter processing: an fmri study*" investigates the neural correlates of rhythm processing in speech perception. The study indicated a function of the SMA and insula beyond motor timing in the analysis of acoustic suprasegmental speech characteristics. Furthermore, the study demonstrated a specific task-related function of the right IFG in the processing of accent patterns and showed right secondary auditory cortex involvement in the explicit perception of auditory cues. The third fMRI experiment entitled "*Evidence for a general segmental processing device in the human auditory dorsal stream*" investigates the functional organization of sub-lexical auditory perception with specific respect to auditory spectro-temporal processing in speech and non-speech sounds. This study evidenced the specific involvement of a subdominant left hemispheric fronto-parietal network during the processing of rapidly changing temporal characteristics regardless of the speechness of the sounds. In the fourth experiment entitled "*Electrical brain imaging reveals overlapping activity elicited by temporal cue processing in speech and non-speech sounds*" scalp AEPs in response to CV-syllables and non-speech analogues with varying voice-onset-times (VOT) and noise-onset-times (NOT), were recorded. This study showed that the characteristic AEP waveform in

response to consonant-vowel-syllables can be resembled by non-speech sounds with similar temporal characteristics, and that the primary auditory cortex is asymmetrically activated in favoring the left hemisphere independently of the stimulation used.

The present work was conducted to make a contribution to the ongoing debate on how the brain masters the intractable task of performing auditory speech processing.

In the last years, several models of the neurofunctional base of speech processing have been developed. Scott and Wise (2004) proposed a functional neuroanatomical framework for language perception, assuming two distinct processing pathways involved in speech perception. An anterior processing stream running anteriorlaterally to the primary auditory cortex, which is involved in sound-to-meaning mapping and a posterior stream, located in the temporo-parietal-junction and involved in the mapping of speech sounds to motor representations of articulations (Scott and Wise, 2004). Similarly, Hickok and Poeppel (2004) proposed a large-scale model of the functional anatomy of language (Hickok and Poeppel, 2004). Their framework proposes the distinction between two processing streams; a functional dorsal-ventral partitioning. The ventral stream, which projects ventro-laterally and involves the STS and the posterior MTG and ITG is involved in sound-to-meaning mapping and the dorsal stream, which projects dorso-posteriorly toward the inferior parietal lobe and frontal regions is involved in sound-to-articulartory-based representations. This dorsal pathway is supposed to play an important role in sub-lexical speech perception (Hickok and Poeppel, 2000). Poeppel and Hickok hypothesized that the performance of sub-lexical tasks will specifically engage the dorsal processing path connecting left inferior frontal and left inferior parietal cortex as part of a network for audio-motor integration. As suggested by Hickok et al., (2000) this fronto-parietal network predominantly in the left hemisphere might be tuned during development because attention to phonetic segments provides the basis for the learning process of speech

articulation and functions as an interface of auditory and articulatory representations of speech.

Inherent in all these models of speech perception is an assumption of a functional asymmetry between the two cortical hemispheres, in particular the temporal lobes. These complementary roles of the left and right temporal cortex in auditory processing have traditionally been investigated by neuropsychological tests in brain-lesioned patients and more recently, with functional and structural brain imaging methods. The majority of these studies have confirmed the predominant role of the left hemisphere in speech processing and contrary, a predominant role of the right hemisphere in music processing. Nevertheless this converging results and interpretations, the exact role of left and right auditory cortex function in speech and music processing are still a matter of great debate. Several theories concerning lateralized auditory processing have been postulated. It has been proposed that speech related left hemisphere lateralization arises as a result of the linguistic relevance and intelligibility of the input (Scott *et al.,* 2000). Alternatively, it has been argued that the hemispheric specialization exists as a result of asymmetries in basic auditory processing (Efron, 1963; Tallal and Piercy, 1973). More recently, the "spectro-temporal trade-off" hypothesis of Zatorre and Belin (2001) stated that the left hemisphere has finer temporal resolution, whereas the right hemisphere has finer spectral resolution (Zatorre and Belin, 2001). These differences might originate in structural asymmetries between the hemispheres. Likewise, the hypothesis by Poeppel (2001, 2003) suggested that the hemispheric asymmetry arises due to differences in the time window preferentially processed by the hemispheres (Poeppel, 2001; Poeppel, 2003). This asymmetric sampling in time (AST) model posits that the left hemisphere preferentially extracts information over shorter temporal integration windows and the right hemisphere over longer time windows.

The converging results of the present work demonstrate that the human auditory system of the left cortical hemisphere preferentially processes rapidly changing acoustic information independently of the

linguistic content of the sounds. We identified the primary auditory cortex (Heschl's gyrus, HG) and the posterior secondary auditory cortex (planum temporale, PT) as the core regions responsible for the basic analysis of acoustic parameters in the temporal domain. By using varying research methods with complementary advantages, namely fMRI and EEG, we consistently showed a specific involvement of the left HG and PT during the perception of syllables and non-speech sounds with analogous temporal characteristics. Thus, we clearly evidenced the existence of a functional lateralization for the processing of rapidly changing temporal information. Furthermore, by showing a similar neurophysiological reaction in response to speech and non-speech sounds by means of overlapping activations and by means of homogeneous AEP waveform, we suggest that in terms of auditory cortex functions, rapid temporal information processing contributes as the common base of functional lateralization in auditory speech perception. This suggestion has strong implications for the discussion whether brain regions are specifically devoted to processing speech or whether they are specialized for processing several features of sounds in general. This question is closely related to the ongoing debate to what extent speech receives a full auditory analysis before it is processed as speech or whether speech is processed differently at a very early stage. It has been argued that the complexity of speech needs a unique processing mode with unique neural foundations (Liberman and Mattingly, 1985; Liberman and Whalen, 2000). The identification of areas supporting both, acoustic temporal processing during phonetic as well as non-speech perception leads to a better understanding of the issue of generality of speech processing, and provides support for the origin and development of several speech related disorders (e.g. dyslexia). Furthermore, by investigating top-down influences on auditory discrimination tasks, we showed that in an initial stage auditory sub-lexical processing of speech and non-speech sounds is performed by bilateral, but left dominant, superior temporal cortex, which is not affected by top-down manipulations. Additionally, we demonstrated that the explicit processing of temporal information in both speech and non-speech sounds, which requires the extraction of segmental information specifically engaged a

left hemispheric fronto-parietal cortical network. Thus, the present result further evidenced the existence of a dorsal processing pathway in the human brain connecting the left inferior frontal and the left inferior parietal cortex. We demonstrated an involvement of this processing stream in the analysis of sub-lexical auditory speech perception, but also in the processing of non-speech sounds with analogous spectro-temporal characteristics. Thus, the dorsal processing pathway in the auditory domain seems to be specifically dedicated to a general acoustic analysis of sounds within a short time frame rather than displaying specificity for sub-lexical speech processing.

5.3 Reference List

Efron, R., 1963. Temporal Perception, Aphasia and D'ej'a vu. Brain 86, 403-424.

Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci. 4, 131-138.

Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67-99.

Liberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised. Cognition 21, 1-36.

Liberman, A.M., Whalen, D.H., 2000. On the relation of speech to language. Trends Cogn Sci. 4, 187-196.

Poeppel, D., 2001. Pure word deafness and the bilateral processing of the speech code. Cognitive Science 25, 679-693.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication 41, 245-255.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 Pt 12, 2400-2406.

Scott, S.K., Wise, R.J., 2004. The functional neuroanatomy of prelexical processing in speech perception. Cognition 92, 13-45.

Tallal, P., Piercy, M., 1973. Defects of non-verbal auditory perception in children with developmental aphasia. Nature 241, 468-469.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb.Cortex 11, 946-953.

6. Curriculum vitae

Education and Qualifications

2007	University of Zurich, Switzerland, Department of Psychology Ph.D. "Psychology/Neuropsychology" Thesis: "Temporal information processing as a foundation of speech perception", (summa cum laude) Supervisors: Lutz Jäncke, Ph.D
Since 2004	University of Zurich, Switzerland Neuroscience Centre Zurich, International Ph.D. Program in Neuroscience, Supervisor: Fred W. Mast, Ph.D., Lutz Jäncke, Ph.D
2003	University of Magdeburg, Germany, Master of Science (M.S.) Clinical and Experimental Psychology, Master thesis: "Die Detektion kurzer Lücken in Schallsignalen als grundlegende Komponente der Sprachverarbeitung – eine fMRT - Untersuchung" ("Rapid auditory perception as the foundation of speech processing – an fMRI study"), (grade 1.0) Supervisor: Lutz Jäncke, Ph.D., Kirsten Jordan, Ph.D.

Fellowships and Awards

2005	Travel Award of Swiss Society for Neuroscience (SSN) in support attendance at the Organization for Human Brain Mapping's Annual Meeting in Toronto (Canada), June 12-16, 2005.
2000 – 2003	Fellowship from the German National Merit Foundation (Studienstiftung des Deutschen Volkes)

Research Experience

Since 2004	Research Fellow: University of Zurich, Department of Cognitive Neuroscience and Neuropsychology, Zurich, Switzerland, Supervisors: Fred W. Mast, Ph.D., Lutz Jäncke, Ph.D.
2003	Research Fellow: University of Magdeburg, Klinik für Kinder- und Jugendpsychiatrie, Germany, Supervisor: Kerstin Kraul, Ph.D.
05/03 – 08/03	Internship: Max-Planck-Institute for Cognitive Neuroscience, Leipzig, Germany, Responsibilities: Evaluation of statistical methods for fMRI data analysis. Supervisor: Arturo Hernandez, Ph.D.
07/01 – 11/01	Internship: Harvard Medical School, Beth Israel Deaconess Medical Centre, Department of Neurology, Boston, MA, Responsibilities: Gaining experience in neuroscientific experimental settings, data acquisition, and statistical analysis. Supervisor: Gottfried Schlaug, MD, Ph.D.
07/00 – 09/00	Internship: University of Magdeburg, Department of Neurology II, Germany Responsibilities: Gaining experience in fMRI data acquisition and analysis. Supervisor: Claus Tempelmann, Ph.D.
03/99 – 09/03	Research Assistant: University of Magdeburg, Department of General Psychology, Germany

Member of Scientific Associations

CNS	Cognitive Neuroscience Society
OHBM	Organization for Human Brain Mapping
SSN	Swiss Society for Neuroscience
ZNZ	Neuroscience Center Zurich

Own Publication

Peer - reviewed publications:

Zaehle, T., Jancke, L., & Meyer, M. (under Review). Electrical brain imaging evidences left auditory cortex involvement in speech and non-speech discrimination based on temporal features.

Zaehle, T., Geiser, E., Alter, K., Jancke, L., & Meyer, M. (under Review). Evidence for a general segmental processing device in the human auditory dorsal processing stream.

Geiser, E., **Zaehle, T.**, Jancke, L., & Meyer, M. (accepted for publication). The neural correlate of speech rhythm as evidenced by meter processing: an fmri study. *Journal of Cognitive Neuroscience*.

Zaehle, T., Clapp, W.C., Hamm, J.P., Meyer, M., & Kirk, I.J. (in press) Induction of LTP-like Changes in Human Auditory Cortex by Rapid Auditory Stimulation: An fMRI Study. *Restorative Neurology and Neuroscience*.

Zaehle, T., Schmidt, C. F., Meyer, M., Baumann, S., Baltes, C., Boesiger, P. & Jancke, L. (2007). Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. *Neuroimage*. 37, 1195-1204.

Zaehle, T., Jordan, K., Wustenberg, T., Baudewig, J., Dechet, P. & Mast, F.W. (2007). The Neural Basis of the Egocentric and Allocentric Spatial Frame of Reference. *Brain Research* 137(1), 92-103.

Schmidt, C. F.*, **Zaehle, T.***, Meyer, M., Geiser, E., Boesiger, P. & Jancke, L. (2007) (* shared first authorship). Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task. *Human Brain Mapping*, (Epub ahead of print).

Clapp,W.C., **Zaehle,T.**, Lutz,K., Marcar,V.L., Kirk,I.J., Hamm,J.P., Teyler,T.J., Corballis,M.C., and Jancke,L. (2005). Effects of long-term potentiation in the human visual cortex: a functional magnetic resonance imaging study. *Neuroreport* 16, 1977-1980.

Meyer, M., **Zaehle, T.**, Gountouna, V.-E., Barron, A., Jancke, L., and Turk, A. (2005) Spectral-temporal processing during speech perception involves left posterior auditory cortex. *Neuroreport* . 16, 1985-1989.

Zaehle, T., Wüstenberg, T., Meyer, M. and Jancke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing – a sparse temporal sampling fMRI study. *European Journal of Neuroscience*, 20, 2447-2456.

Gaab, N., Gaser, C., **Zaehle, T.**, Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory-an fMRI study with sparse temporal sampling. *Neuroimage*, 19, 1417-1426.

Book Chapter:

Mast, F. & **Zaehle, T.** (2007). Spatial reference frames used in mental imagery tasks. In J. Rieser, D. Ashmead, F. Ebner & A. Corn (Eds.), Blindness and brain plasticity in navigation and object perception. Mahwah, NJ: Lawrence Erlbaum Associates.

Published Abstracts:

Zaehle, T., Meyer, M. & Jancke, L. (2006). Spectro-temporal processing during speech perception. Journal of the International Neuropsychological Society 12 (Suppl. 2), 51

Zaehle, T., Jordan, K., Wüstenberg, T., v. Steinbüchel, N., & Mast, F.W. (2006). The Neural Basis of the Egocentric and Allocentric Spatial Frame of Reference. Journal of the International Neuropsychological Society 12 (Suppl. 2), 51

Gaab,N., **Zaehle,T.**, Gaser,C., Chen,Y., and Schlaug,G. (2002). The functional anatomy of pitch memory: An fMRI-study with sparse temporal sampling. Neurology 58, A264-A265.

Gaab,N., **Zaehle,T.**, Gaser,C., Chen,Y., and Schlaug,G. (2002). The role of posterior Perisylvian regions in pitch memory: An fMRI-study with sparse temporal sampling. Journal of Cognitive Neuroscience 169.

Wüstenberg T, Fesl G, **Zahle T**, Palitzsch M, Perckhofer M et al. (2001). Auditory perception of temporal-spatial order – a fMRI-study. Neuroimage 13(Suppl):961.

Wüstenberg, T., Rüsseler, J., Fesl, G., **Zähle, T.**, Palitzsch, M., Perkhofer, M., Heinze, H-J., Yousry, T.A., Jäncke, L. & v. Steinbüchel, N. (2001). Auditory perception of temporal-spatial order before and after training: Evidence from psychophysics, ERPs and fMRI. Behavioural Pharmacology 12 (Suppl. 1), 111.

Congress Posters / Abstracts

Zaehle, T., Jancke, L. & Meyer, M. (2007). Auditory Perception of Spectral and Temporal Variations in Speech and Nonspeech Sounds – A Sparse fMRI Study.
Poster presented at: the Cognitive Neuroscience Society Annual Meeting (CNS), New York, NY, May 2007.

Zaehle, T., Meyer, M. & Jancke, L. (2006). Spectro-temporal processing during speech perception.
Poster presented at: the Midyear Meeting of the International Neuropsychological Society (INS), Zurich, Switzerland, July 2006.

Zaehle, T., Jordan, K., Wüstenberg, T., v. Steinbüchel, N., & Mast, F.W. (2006). The Neural Basis of the Egocentric and Allocentric Spatial Frame of Reference.
Poster presented at: the Midyear Meeting of the International Neuropsychological Society (INS), Zurich, Switzerland, July 2006.

Zaehle, T., Meyer, M. & Jancke, L. (2006). Spectro-temporal processing during speech perception involves left posterior auditory cortex.
Poster presented at: 12th Annual Meeting of the Organization for Human Brain Mapping, Florence, Italy, June 2006.

Clapp, W.C., **Zaehle, T.**, Lutz, K, Marcar, V L., Kirk, I.J., Hamm, J.P., Teyler, T.J., Corballis, M.C. & Jancke, L. (2006). Induction of long-term potentiation in human visual cortex by rapid visual stimulation: an fMRI-study.

Poster presented at: 12th Annual Meeting of the Organization for Human Brain Mapping, Florence, Italy, June 2006.

Zaehle, T., Jordan, K., Wüstenberg, T., v. Steinbüchel, N., & Mast, F.W. (2005). Distinct neural basis for computing the egocentric and allocentric spatial frame of reference. Poster presented at: 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto, Canada, June 2005.

Wüstenberg, T., **Zaehle, T.**, Meyer, M., Strasburger, H., & v. Steinbüchel, N. (2005). Rapid temporal processing in auditory cortex: two comparative fMRI-experiments. Poster presented at: 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto, Canada, June 2005.

Jola, C., **Zaehle, T.**, Lobmaier, J., Dietrich, Th. & Mast, F. (2004). Mental object rotation vs. egocentric body transformation: A question of rotation axis? Interdisciplinary College (IK) 2004, Günne at Lake Möhne, Germany, 05-12. Mars 2004.

Gaab, N., **Zähle, T.**, Gaser, C., Chen, Y. & Schlaug, G. (2002). The role of posterior perisylvian regions in pitch memory: an fMRI-study with sparse temporal sampling. Poster presented at: "54th Annual Meeting of the American Academy of Neurology" in Denver, CO, April 2002.

Gaab, N., **Zähle, T.**, Gaser, C., Chen, Y. & Schlaug, G. (2002). The role of posterior perisylvian regions in pitch memory: an fMRI-study with sparse temporal sampling. Poster presented at: "Ninth Annual Meeting of the Cognitive Neuroscience Society" in San Francisco, CA, April 2002.

Wüstenberg, T., Fesl, G., **Zähle, T.**, Palitzsch, M., Perkhofer, M., Heinze, H-J., Yoasry, T.A., Jäncke, L. & v. Steinbüchel, N. (2001). Auditory perception of temporal-spatial order – an fMRI study. Poster presented at: "7th International Conference on Functional Mapping of the Human Brain" in Brighton, UK, June 2001.

Wüstenberg, T., Rüsseler, J., **Zähle, T.**, Fesl, G., Wolf, C., Palitzsch, M., Jäncke, L. and v. Steinbüchel, N. (2001). Auditory perception of temporal-spatial order before and

after training: evidence from psychophysics, ERPs and fMRI.
Poster presented at: “Meeting of the European Brain and Behaviour Society (EBBS) and European Behavioural Pharmacology Society”, in Marseille, France, September 2001.